BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

REMARKS

Claims 121-145 are pending and being examined. Claims 121, 122 and 129-131 are amended herein.

Amendment of the claims and specification do not involve new matter and are supported by the specification as originally filed. Entry of these amendments is respectfully requested.

Support for the amended specification at page 10, lines 14-21 can be found in the originally-filed specification of the subject application at page 25, lines 9-19; and Figure 17. Also this amendment includes correction of a typographical error.

Support for the amended specification at page 10, lines 23-31 and continuing at page 11, lines 1-3 can be found in the originally-filed specification of the subject application at Figure 18.

Support for the amended specification at page 11, lines 5-28 can be found in the originally-filed specification of the subject application at page 87, lines 4-25; and Figure 19.

Support for the amended specification at page 11, lines 30-31 and continuing at page 12, lines 1-6 can be found in the originally-filed specification of the subject application at page 87, lines 27-31 and continuing at page 88, lines 1-13; and Figure 20.

The amendment of the specification at page 12, lines 14-24 is merely to correct a typographical error.

The amendment of the specification at page 12, lines 26-31 and continuing at page 13, lines 1-18 is merely to correct a typographical error.

Support for amended claims 121 and 122 can be found in the originally-filed specification of the subject application at page 21, lines 24-29; page 22, lines 11-17; page 22, lines 26-31; page 23, lines 1-2 and lines 11-16 and lines 19-30; page 24, lines 1-9; and Figure 17A.

The amendment to claims 129-130 are merely to add sequence listing indicators. Support for amended claims 129-130 can be found in the originally-filed specification of the subject application at page 97, lines 2-27.

The amendment to claims 129-130 are merely to add sequence listing indicators. Support for amended claim 131 can be found in the originally-filed specification of the subject application at page 14, line 27; page 96, lines 1-3; and Figure 27.

Oath/Declaration:

In the Office Action, the Patent Office deemed the Declaration previously submitted was defective because non-initialed and/or non-dated alterations have been made.

Applicants are currently in the process of executing the amended Combined Declarations and Power of Attorney. Executed documents will be forwarded to the Patent Office once they are availble.

Drawings:

In the Office Action, the Patent Office objects to the drawings because Figure 17 contains parts A and B. In response, Applicants have amended the Brief Description Figure 17 to include parts A and B.

The Brief Description of Figure 18 describes Ad5/11 which does not appear in the figure. In response, Applicants have amended the Brief Description of 18 to delete "Ad5/11".

Also, Figures 19-20 contain parts C and D which are not described in the Brief Description of the Drawings. In response, Applicants have amended the Brief Description Figures 19 and 20 to include parts A and B.

Figures 22 and 23 refer to Figure 8 text for further descriptions of the details, but Figure 8 does not correspond to the details indicated in Figures 22 and 23. In response, Applicants have amended the specification to delete the references to Figure 8.

Figure 29 does not contain parts A-D which are described in the Brief Description of the Drawings. In response, Applicants submit herein a corrected drawing of Figure 29 showing parts A-D.

Additionally, Applicants submit herein a complete set of formal drawings of Figures 1-29 (EXHIBIT 1). The formal drawings do not contain new matter. Applicants respectfully request their entry.

Sequence Compliance:

In the Office Action, the Patent Office requires Applicants to submit a substitute sequence listing, computer readable form, and a statement, because Figure 27 and claims

129-131 contain sequences which are not included in a previously submitted sequence listing.

In response, Applicants submit herein a substitute sequence listing in paper form (EXHIBIT 2) and computer readable form, and a Declaration under 37 C.F.R. §1.821(f) (EXHIBIT 3).

Applicants' Invention:

Applicants teach recombinant, double-stranded, hybrid Ad-AAV vectors, comprising a parallel and an anti-parallel DNA strand, where the parallel strand comprises the following elements: a left and right Ad.ITR sequence; an Ad packaging sequence; a first and second AAV.ITR; a first and second IR sequence; a heterologous promoter; a foreign gene sequence; and a gene sequence that mediates replication of the adenovirus in a transduced cell. The anti-parallel strand of the claimed vectors comprises a nucleotide sequence encoding a modified adenovirus fiber protein which alters the tropism of the adenovirus vector.

REJECTION UNDER 35 U.S.C. §112, FIRST PARAGRAPH:

Rejection of Claim 131:

In the Office Action, the Patent Office issued a new matter rejection. The Patent Office alleges claim 131 contains subject matter which is not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

The Patent Office alleges the peptide ligands having amino acid sequences LNFCSFC and LNGCGXXXXXXXXXXGC are not supported in the specification at page 14, line

27, and Figure 27. Further, the Patent Office alleges the specification does not teach a use of these peptide ligands.

Applicants respectfully disagree. The specification teaches various embodiments of heterologous peptides replacing a GH or HI loop region in order to retarget the gutless vector to a desired cell type. Support can be found in the originally-filed application at page 22, lines 14-17; and page 27, lines 16-24.

Applicants also note that the amino acid sequence in claim 131(c) contains a typographical error. Applicants now amend claim 131(c) to recite the sequence LNGCGSGC. Support for amended claim 131c can be found in the originally-filed specification at page 14, line 27, and Figure 27 (please see "adeno 5 GH cys 2"). Further support can be found page 27, lines 16-24.

Applicants contend that the amino acid sequence in claim 131d, namely LNGCGXXXXXXXXXXGC, is indeed supported in the originally-filed specification at page 14, line 27 and at Figure 27. In particular, Figure 27 shows a GH peptide having this sequence (please see "adeno 5 GH peptides"). Further support can be found page 27, lines 12-20.

Rejection of Claims 121-145:

Inverted Repeat Sequences:

In the Office Action, the Patent Office rejects claims 121-145 under 35 U.S.C. 112, first paragraph, because the specification supports the claimed adenoviral vector having pairs of Ad ITR's and AAV ITRs, but the specification allegedly does not support a third pair of inverted repeat sequences.

In response, Applicants respectfully disagree. Applicants contend that the originally-filed claims and specification provide support for an adenoviral vector having the third pair of inverted repeat (IR) sequences.

For example, pending claims 121 and 122 contain the same subject matter of originally-filed claims 16, 26, 27, 28 and 33.

Claim 16 recites a double-stranded Ad vector including a first strand (i.e., parallel strand) comprising: a left and right Ad-ITR; an Ad packaging sequence; a transgene cassette; and at least one Ad sequence which directs adenoviral replication. The second strand (i.e., anti-parallel strand) encodes an adenoviral fiber protein that permits targeting of the Ad vector into the host cell of interest.

In an embodiment of claim 16, originally-filed claims 26 and 27 recite the transgene cassette comprises: a left and right ITR; a polyadenylation sequence; a transgene sequence; and a promoter sequence.

In a further embodiment of the left and right ITR of claims 26 and 27, originally-filed claim 28 recites adenoviral-associated ITR's.

And in a further embodiment of claims 26 and 27, originally-filed claim 33 recites the transgene cassette further comprising an inverted repeat sequence.

Thus, in pending claims 121 and 122, support for inverted repeat sequences can be found in originally-filed claims 16, 26, 27, 28 and 33.

In the originally-filed specification, support for the inverted repeat sequences can be found at: Figure 17A; the Brief Description of Figures at page 10, lines 14-21; and at Example II G at page 86, lines 10-20.

Source of the Inverted Repeat Sequences:

In the Office Action, the Patent Office also alleges Applicants provide no guidance for sequences that are meant to constitute the inverted repeat sequences in claim 121 and 122, steps (d) and (g).

Applicants respectfully disagree. Applicants contend the originally-filed specification discloses the inverted repeat sequences which "mediate predictable genomic rearrangements resulting in a gutless vector genome devoid of all viral genes" (page 23, lines 19-21).

Further, the originally-filed specification discloses the inverted repeat sequence is a 1.2kb inverted homology element (page 10, lines 16-17).

One embodiment of the inverted repeat sequence is the 1.2kb fragment from a chicken globin HS-4 element. Support for this embodiment can be found in Steinwaerder, et al., 1999, *J. Virology* 73:9303-9313, which is incorporated by reference (please see the originally-filed specification at page 32, lines 23-24) (a copy of Steinwaerder is attached as EXHIBIT 4).

Steinwaerder, et al., discloses producing gutless adenoviral vectors comprising the 1.2kb chicken globin HS-4 element which includes inverted repeat sequences (see Steinwaerder, et al., at: Abstract, lines 4-6; also page 9304, right column, lines 5-9 under "Results"). Steinwaerder, et al., reports that the chicken globin HS-4 element mediates predictable genomic rearrangements when inserted into the E1 region of an adenoviral vector.

Additionally, the nucleotide sequence of the HS-4 element was known to those skilled in the art at Genbank, accession no. U78775 (submitted November 18, 1996 to National Institute of Health; see also Steinwaerder, et al., at: page 9304, left column, lines 1-4 after "Materials and Methods").

Applicants also contend it was well known to the skilled artisan that other inverted repeat sequences mediate DNA recombination processes, including: bacterial transposons Tn5 and Tn10 (EXHIBIT 5 is Gordenin, et al., September 1993, Molecular and Cellular Biology 13:5315-5322); sequences contained in the left and right terminal ends of Ad5 genome (EXHIBIT 6 is Chartier, et al., July 1996 Journal of Virology 70: 4805-4810); yeast ARM's or DNA at-risk-motifs (EXHIBIT 7 is Gordenin and Resnick May 1998 Mutat. Res. 400:45-58); transposable elements from insects (EXHIBIT 8 is O'Brochta and Atkinson, Sept-Oct 1996 Insect Biochem Mol Biol 26:739-753); and direct and inverted tandem repeat sequences from human (EXHIBIT 9 is Lambert, et al., Apr 1999 Mutat. Res. 433:159-168).

The Modified Fiber Protein:

In the Office Action, the Patent Office states that the specification does not teach how to modify the second strand to encode a modified fiber (Office Action, page 6, line 7-8; page 6 line 22- page 7, line1).

Applicants respectfully disagree. The specification discloses many embodiments of modified adenoviral fiber proteins.

For example, the specification discloses heterologous peptide sequences comprising an Ad5 fiber tail domain and an Ad35 fiber shaft and knob domain (page 18, lines 9-16), or an Ad5 fiber tail domain and Ad11 fiber shaft domain (page 26, lines 27-30), or fiber

proteins from Ad 3, 7, 9, 11, and 35 (page 26, lines 15-20). The resulting Ad5/35 and Ad5/11 fiber proteins showed broader tropism (See Table I, below).

Applicants also provide an extensive list of various adenoviral serotypes which can be used by the skilled artisan to retarget the host cell of interest (see Appendix I in the specification as originally filed at pages 107-113).

Applicants teach how to make a hybrid Ad-AAV vector comprising a nucleotide sequence encoding a modified adenoviral fiber protein. Support can be found in the originally-filed specification at Example II, Section G, at page 83, lines 14-31; page 84, lines 1-4; page 85, lines 15-20; and Figures 16A and 18.

Other examples include replacing the G-H or H-I loop with heterologous peptide sequences such as the RI or RII protein from malaria circumsporozoite surface protein (page 27, lines 28-30; page 28, lines 6-12; and Example II J at page 96, line 31 and continuing at page 97, lines 1-27)

In other embodiments, the G-H or H-I loop is replaced the two peptides such as the RGD-4 peptide and a peptide that binds matrix metalloproteinase (page 98, lines 1-10).

Yet other embodiments include heterologous peptides depicted in Figure 27.

Support for Ad.AAV hybrid vectors comprising nucleotide sequences encoding modified fiber proteins that alter the tropism of the vector can be found in the originally-filed specification in the Figures, as shown in Table 1 below.

In particular, the data in Figure 25 shows virions encoded by the Ad5L vector attach, internalize and transduce 293 cells with high efficiency but to K562 cells with low efficiency. The virions encoded by the Ad5/35L and Ad5/35S vectors attach, internalize

and transduce both 293 and K562 cells with high efficiency. With this data Applicants teach how to make and use an Ad.AAV vector encoding a modified fiber protein that alters tropism of the encoded virion.

TABLE I

Figures:	Depicts:	Ad vectors/Cells:
Figure 12	Tropism (e.g., attachment and internalization) of non-modified Ad virions to cells	Ad 3, 4, 5, 9, 35 and 41 CHO (ovary) HeLa (cervical) K-562 (hematopoietic) CD34+ (bone marrow)
Figure 13	Tropism (e.g., attachment and internalization) of non-modified Ad virions to cells	Ad 3, 5, 9, 35 and 41 CHO (ovary) HeLa (cervical) 293 (kidney)
Figure 14	Tropism (e.g., attachment and internalization) of non-modified Ad virions to cells	Ad 3, 5, 9, 35 and 41 K-562 (hematopoietic) CD34+ (bone marrow)
Figure 16	Schematic showing vector construct of a serotype hybrid Ad construct having Ad5 tail and Ad35 knob and short shaft: "Ad5-GFP-F35" see also page 9, lines 21-31 and page 10, lines 1-12	Ad5/Ad35 hybrid
Figure 18	Schematic showing vector construct of a serotype hybrid Ad construct having Ad5 tail and Ad35 knob and short shaft: "Ad5-GFP-F35"	Ad5/Ad35 hybrid
Figure 19	Cross competition data showing tropism (e.g., attachment and internalization) of Ad5-GFP-F35 virions to K-562 cells	Ad5/Ad35 hybrid K-562 (hematopoietic)
Figure 20	Cross competition data showing tropism (e.g., attachment and internalization) of Ad5-GFP-F35 virions to K-562 cells	Ad5/Ad35 hybrid K-562 (hematopoietic)
Figure 21	Graph showing transduction (gene transfer) of cells with Ad 5-GFP-F35 vector	Ad5/Ad35 hybrid CD34+ (bone marrow) K-562 (hematopoietic) HeLa (cervical)

Figures:	Depicts:	Ad vectors/Cells:
Figure 22	FACS analysis of CD34+ cells transduced with Ad5-GFP-F35, and then sorted for transduced cells that express GFP (from transduction)	Ad5/Ad35 CD34+ (bone marrow)
Figure 23	FACS analysis of CD34+ cells transduced with Ad5-GFP-F35, and then sorted for transduced cells that express GFP (from transduction) and also express or CD117+ (a.k.a: c-kit)	Ad5/Ad35 CD34 and CD117 (precursor bone marrow cells, e.g., stem cells)
Figure 25	Upper: schematic depicting hybrid Ad virions having Ad5 knob with stem and tail of Ad9 or Ad35 Lower: table showing tropism data (e.g., attachment, internalization, and transduction data) of hybrid virions	Ad5/Ad9 Ad5/Ad35 293 (kidney, CAR+) K-562 (hematopoietic, little CAR) Y79 (CAR+)
Figure 27	Schematic depicting Ad5, wild-type sequence of GH loop, and also showing 4 variants (mutant sequences) or GH loop	Sequences of mutant GH loop
Figure 28	Tropism (e.g., attachment and internalization) of non-modified Ad virions to cells	Ad 3, 5, 9, 35 and 41 REVC (vascular endothelial) JAWSII (dendritic) MCF7 (breast cancer) 8714 (T lymphocyte)

REJECTION UNDER 35 U.S.C. §112, SECOND PARAGRAPH:

In the Office Action, the Patent Office rejects claims 121-145 under 35 U.S.C. §112, second paragraph because the terms "first strand" and "second strand" are unclear. In particular, the Patent Office states that it is not clear whether the "first and second strands" correspond to the r and l strands of the double-stranded adenovirus or to the segments of DNA that are each double-stranded.

In response, Applicants contend the terms "first strand" and "second strand" refer to the inventive adenoviral vector comprising a first single strand of DNA which also is termed a parallel strand, and a second single strand of DNA strand which is anti-parallel to the first DNA strand.

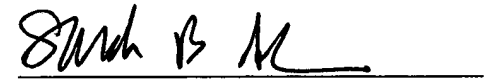
Applicants have amended claims 121 and 122 to recite parallel and anti-parallel strands. Support for this amendment can be found in the originally-filed specification at page 19, lines 24-30 and Figure 1A which shows a double-stranded vector each having a 5' end.

CONCLUSION

If a telephone interview would be of assistance in advancing prosecution of the subject application, Applicants' undersigned attorney invites the Examiner to telephone her at the number provided below.

The Patent Office is authorized to charge a one-month extension of time fee under 37 C.F.R. §1.17(a)(1) of $55.00 to Deposit Account No. 50-0306. No fee, other than the fee for a one-month extension of time, is deemed necessary in connection with the filing of this Communication. However, if any additional fee is necessary, the Patent Office is authorized to charge any additional fee to Deposit Account No. 50-0306.

Respectfully submitted,



Sarah B. Adriano
Registration No. 34,470
Mandel & Adriano
55 So. Lake Ave., Suite 710
Pasadena, California 91101
626/395-7801
Customer No: 26,941

JOURNAL OF VIROLOGY, Nov. 1999, p. 9303–9313
0022-538X/99/$04.00+0

Vol. 73, No. 11

Generation of Adenovirus Vectors Devoid of All Viral Genes by Recombination between Inverted Repeats

DIRK S. STEINWAERDER, CHERYL A. CARLSON, AND ANDRÉ LIEBER*

Division of Medical Genetics, University of Washington, Seattle, Washington 98195

Received 3 June 1999/Accepted 15 August 1999

Direct or inverse repeated sequences are important functional features of prokaryotic and eukaryotic genomes. Considering the unique mechanism, involving single-stranded genomic intermediates, by which adenovirus (Ad) replicates its genome, we investigated whether repetitive homologous sequences inserted into E1-deleted adenoviral vectors would affect replication of viral DNA. In these studies we found that inverted repeats (IRs) inserted into the E1 region could mediate predictable genomic rearrangements, resulting in vector genomes devoid of all viral genes. These genomes (termed ΔAd.IR) contained only the transgene cassette flanked on both sides by precisely duplicated IRs, Ad packaging signals, and Ad inverted terminal repeat sequences. Generation of ΔAd.IR genomes could also be achieved by coinfecting two viruses, each providing one inverse homology element. The formation of ΔAd.IR genomes required Ad DNA replication and appeared to involve recombination between the homologous inverted sequences. The formation of ΔAd.IR genomes did not depend on the sequence within or adjacent to the inverted repeat elements. The small ΔAd.IR vector genomes were efficiently packaged into functional Ad particles. All functions for ΔAd.IR replication and packaging were provided by the full-length genome amplified in the same cell. ΔAd.IR vectors were produced at a yield of ~10^4 particles per cell, which could be separated from virions with full-length genomes based on their lighter buoyant density. ΔAd.IR vectors infected cultured cells with the same efficiency as first-generation vectors; however, transgene expression was only transient due to the instability of deleted genomes within transduced cells. The finding that IRs present within Ad vector genomes can mediate precise genetic rearrangements has important implications for the development of new vectors for gene therapy approaches.

The starting point for the presented study was an observation made with first-generation adenovirus (Ad) vectors that contained fragments of Ad5 DNA, specifically the va_I (23) or the precursor to the terminal protein (pTP) (26) genes, inserted into the E1 region. The presence of these sequences in addition to the corresponding endogenous gene resulted in the appearance of two viral bands with different buoyant density in CsCl gradients after ultracentrifugation of lysates from infected 293 cells. This phenomenon was interesting, considering the unique mechanism by which the adenovirus replicates and the functional potential of repetitive sequences to mediate genetic rearrangements.

The genomes of Ad2 and Ad5 are double-stranded, linear DNA molecules, approximately 35 kb in length with an inverted terminal repeat sequence (ITR) of 102 bp on each end. Numerous studies in cell-free systems and in infected cells have established that Ad DNA replication takes place in two steps (reviewed in references 2 and 37). In the first stage, DNA synthesis is initiated by pTP. pTP binds as a heterodimer with the Ad polymerase (Pol) to specific sites within the ITRs. Ad DNA replication begins at both ends of the linear genome, resulting in a daughter strand that is synthesized in the 5′ to 3′ direction, displacing the parental strand with the same polarity. Three nonexclusive mechanisms are proposed for the second step, the replication of the displaced parental strand. (i) Displaced single strands can form partial duplexes by base pairing of the ITRs on which a second round of DNA synthesis may be initiated (22, 36). (ii) When two oppositely moving displacement forks meet, the two parental strands can no longer be held together and therefore separate, resulting in partially duplex and partially single-stranded molecules; the synthesis is then completed on the displaced parental strand (21). (iii) Displaced strands, with opposite polarity resulting from initiation at two different molecular ends, can renature to form a double-stranded daughter molecule (37). Elongation of DNA synthesis requires only DNA binding protein (DBP) and Pol. With 20 to 30 bp being synthesized per second, Ad elongation is relatively slow compared to that in the eukaryotic replication systems (which synthesize ~500 bp/s). DBP may stabilize the formation of the panhandle structure and the interstrand renaturation process (39).

Repetitive sequences are a common feature of prokaryotic and eukaryotic genomes. Direct repeats (DR) and inverted repeats (IR) are associated with DNA recombination processes (5, 20, 29). Furthermore, it is thought that IR-induced DNA secondary structures cause pausing of replication by DNA polymerases and reverse transcriptases, resulting in genetic alterations (1, 7, 12, 13, 19, 38).

The unique Ad replication strategy, involving single-stranded replication intermediates, prompted us to investigate in detail whether repetitive homologous sequences inserted into the Ad vector genome would affect replication of viral DNA or whether it would induce genomic rearrangements. In these studies, we have found that, as a result of the replication of E1-deleted Ad vectors containing IR flanking a transgene cassette, a small viral genome is efficiently formed and packaged. These genomes were devoid of all Ad genes. Particles containing this small genome could be separated from virions with full-length genomes by ultracentrifugation in CsCl gradients. In addition to having interesting virological aspects, this finding has practical importance for Ad vector development.

* Corresponding author. Mailing address: Division of Medical Genetics, Box 357720, University of Washington, Seattle, WA 98195. Phone: (206) 221-3973. Fax: (206) 685-8675. E-mail: lieber00@u.washington.edu.

Ads have a number of properties that make them attractive vehicles for gene transfer. These include highly efficient mechanisms of gene transfer to a large variety of cell types in vivo and the easy production of purified virus at high titers. Highly efficient transduction is mediated by the capsid and core proteins involved in cell attachment and by internalization, endosomal lysis, and nuclear import. Most Ad vectors used for in vivo gene transfer are deleted for E1 genes. The major limitation associated with these E1-deleted vectors has been in short-term expression in vivo, due to the development of immune responses to expressed viral proteins which result in toxicity and viral clearance. In order to overcome some of these problems, Ad vectors have been developed from which almost the entire Ad genome has been deleted. These include vectors with "gutless," almost full-length genomes, which have been shown to mediate stable transgene expression in vivo (32), as well as encapsidated Ad minichromosomes with genomes of ~13 kb, which also have successfully been used for gene transfer in vitro and in vivo (8, 17, 18). Both vector systems require helper viruses and several serial passages for production.

As an application of our finding that IRs can mediate predictable genetic rearrangements within Ad genomes, we demonstrate here the efficient and straightforward production of new vectors representing small Ad genomes devoid of all viral genes, which are packaged into functional Ad capsids.

MATERIALS AND METHODS

Production and characterization of viral vectors. Plasmids. Sequences of the metal responsive element (MRE) and HS-4 element were taken from the plasmids pMRENeo (33) (provided by Richard Palmiter, University of Washington) and pJC5-4 (GenBank accession no. U78775) (a gift from G. Felsenfeld, National Institutes of Health). The human α_1-antitrypsin (hAAT) cDNA linked to a bovine growth hormone (bPA) gene polyadenylation signal was derived from pBShAAT (14).

All Ad shuttle vectors were based on pΔE1sp1a (Microbix, Toronto, Canada). The construction of the vectors Ad.hAATa/b, Ad.Ins1/1a/b, Ad.Ins1/3a/b, Ad.Ins2/1a/b, Ad.Ins2/2a/b, and Ad.IR(G/C) is described elsewhere (25, 35). Vectors containing shortened IRs [Ad.IR(1.0) and Ad.IR(0.5)] were generated by the digestion of pIns2/2 (35) with *Afl*II and *Hin*dIII. The isolated *Afl*II fragment containing shortened insulator sequences and the MREhAAT cassette was blunted (T4 polymerase) and ligated to the *Eco*RV site of pΔE1sp1a [pAd.IR(0.5)]. Partial digestion of pIns2/2 with *Hin*dIII revealed the 4.1-kb fragment containing two shortened insulators flanking the MREhAAT cassette. The fragment was ligated to the *Hin*dIII site of pΔE1sp1a[pAd.IR(1.0)].

For the construction of Ad.IR-bGal, a transgene cassette containing the β-galactosidase (β-Gal) cDNA fused to the simian virus 40 (SV40) polyadenylation signal (SV40pA) was generated. The SV40pA was cut out from pREP4 (Invitrogen, Carlsbad, Calif.) by *Xho*I/*Sal*I digestion and was ligated to the *Xho*I site of pBluescript SK(+) (Stratagene, La Jolla, Calif.), resulting in pBS-SV40pA. A 3.7-kb *Bam*HI fragment containing the bGal cDNA was derived from pCMVbGal (24) and was inserted into the corresponding site of pBS-SV40pA, resulting in pbGal-SV40pA. Ligation of a bGal-SV40pA containing a 4.1-kb *Xba*I/*Kpn*I fragment of pbGal-SV40pA to the corresponding sites of pSLJCb (35) resulted in pJC(1)bGal. pJC(2)bGal was created by insertion of an HS-4 containing a *Spe*I/*Nhe*I fragment derived from pSLJCa (35) into the *Spe*I sites of pJC(1)bGal. In the right orientation, the complete cassette holding two inverted HS-4 elements flanking bGal-SV40pA was cut from pJC(2)bGal by *Spe*I/*Nhe*I digestion and was inserted into the *Xba*I site of pAd.RSV (14), generating pAd.IR-bGal.

Ads. First-generation viruses with the different transgene cassettes incorporated into their E1 regions were generated by recombination of the pΔE1sp1a-derived shuttle plasmids and pJM17 or pBHG10 (Microbix) in 293 cells as previously described (27). For each virus, at least 20 plaques were picked, amplified, and analyzed by restriction digest. Plaques from viruses with the correct genome structure were amplified, CsCl banded, and titered (in PFU per milliliter) as previously described (14, 27). All virus preparations tested negative for replication-competent Ad and bacterial endotoxin (28). Virus was stored at −80°C in a solution containing 10 mM Tris-Cl (pH 7.5), 1 mM $MgCl_2$, and 10% glycerol.

To produce purified ΔAd.IR-1, 293 cells were infected with Ad.2/2a at a multiplicity of infection (MOI) of 10 PFU/cell and were harvested 40 h after infection. Cells were lysed in phosphate-buffered saline by four cycles of freezing and thawing. Lysates were centrifuged to remove cell debris and were digested for 30 min at 37°C with 500-U/ml DNase I and 200-μg/ml RNase A in the presence of 10 mM $MgCl_2$. Five milliliters of lysate was layered on a CsCl step gradient (0.5 ml at 1.5 g/cm^3, 2.5 ml at 1.35 g/cm^3, and 4 ml at 1.25 g/cm^3) and ultracentrifuged for 2 h at 35,000 rpm (rotor SW41). CsCl fractions were collected by puncturing the tube and were analyzed for viral DNA (27) or were subjected to ultracentrifugation at 35,000 rpm for 18 h in an equilibrium gradient with 1.32 g of CsCl per cm^3. The band containing the deleted virus ΔAd.IR-1 was clearly separated (0.5-cm distance) from other banded viral particles containing full-length Ad.Ins2/2a genomes. Fractions containing deleted virus particles were dialyzed against a solution containing 10 mM Tris-Cl (pH 7.5), 1 mM $MgCl_2$, and 10% glycerol and were stored at −80°C. The genome titer of ΔAd.IR-1 preparations was determined based on quantitative Southern analysis of viral DNA purified from viral particles in comparison to different concentrations of a 1.7-kb hAAT-bPA fragment of pBShAAT (for ΔAd.IR-1) according to a protocol previously described (27). Titers were routinely obtained in the range of 3×10^{12} to 8×10^{12} genomes per ml. Assuming one genome is packaged per capsid, the genome titer equals the particle titer. The level of contaminating Ad.Ins2/2a in ΔAd.IR1 preparations was less than 0.1% as determined by Southern analysis, which is consistent with results obtained by plaque assay of 293 cells (fewer than five plaques per 10^6 total genomes).

Primers used for sequencing the ΔAd.IR-1 genome, specific to Ad5 nucleotides (nt) 319 to 338 and 3550 to 3531, were AdF, 5′-TTGTGTTACTCATAG CGCGT, and AdR, 5′-TTCTTTCCCACCCTTAAGCC. The nested primers to obtain the complete sequence of the IR elements in ΔAd.IR-1 were 5′ (nt 552) TGACATTGTTGGTCTGGC and 5′ (nt 947) GAAAAGCTCCAAGATCCC.

Electron microscopy. For examination of viral particles in the transmission electron microscopy studies, CsCl-purified virions were fixed with glutaraldehyde and were stained with uranyl acetate as described previously (27).

Cell culture. SKHEP-1 cells (HTB-52; American Type Culture Collection, Rockville, Md.), an endothelial cell line derived from human liver (10), were grown in high-glucose Dulbecco's modified Eagle medium with 10% fetal calf serum.

Analysis of viral DNA. Lysates from 2×10^5 cells that had developed complete cytopathic effect (CPE) after viral infection or viral material banded in CsCl gradients were digested with pronase (1 mg/ml in a solution containing 10 mM Tris-Cl (pH 7.4), 10 mM EDTA (pH 8.0) and 1% sodium dodecyl sulfate) for 2 h. DNA was extracted with phenol-chloroform and was precipitated in ethanol. DNA samples were then subjected to gel electrophoresis or restriction analysis.

Southern blotting. Cultured cells were washed three times with phosphate-buffered saline before harvesting. For analysis, 10 μg of genomic DNA was digested with restriction endonucleases at 37°C overnight and then electrophoresed in a 0.8% agarose gel and transferred to a nylon membrane (Hybond N$^+$; Amersham, Arlington Heights, Ill.). The blots were hybridized in rapid hybridization buffer (Amersham) with [α-^{32}P]dCTP-labeled DNA probes (>10^8 cpm/μg of DNA). The fragment used for labeling was the 1.7-kb hAAT-bPA fragment of pBShAAT.

hAAT ELISA. hAAT concentrations in cell culture supernatants were determined by enzyme-linked immunosorbent assay (ELISA) as previously described (14).

RESULTS

Considering the unique mechanism by which Ad replicates its genome, involving single-stranded genomic intermediates able to form intra- and intermolecular hybrids, we decided to study in more detail whether repeated sequences inserted into the viral genome would affect replication of viral DNA. To this end, a number of first-generation Ad vectors were used that contained single or repeated copies of a 1.2-kb chicken globin HS-4 element (3) inserted into the E1 region together with a reporter gene cassette (Fig. 1A). These vectors were originally designed for an unrelated study that employed the HS-4 element as an insulator to shield a heterologous, inducible promoter from interference by Ad enhancers (35). Control vectors consisted of the promoterless transgene only (Ad.hAATa) or the transgene expression cassette combined with one HS-4 element (Ad.Ins1/1a and Ad.Ins1/3a). Ad.Ins2/2a and Ad.Ins2/2b contained the HS-4 elements as IRs flanking the reporter gene cassette in leftwards or rightwards orientation, respectively. In Ad.Ins2/1a and Ad.Ins2/1b, the transgene cassette was flanked by HS-4 DRs. In a first screening for abnormal vector replication products, viral DNA was isolated together with chromosomal DNA from infected 293 cells after the development of full CPE and was analyzed by gel electrophoresis (Fig. 1B). The full-length (~35-kb) genome comigrated with fragments of genomic DNA. Interestingly, a small (~5.7-kb) band appeared in DNA samples isolated from cells

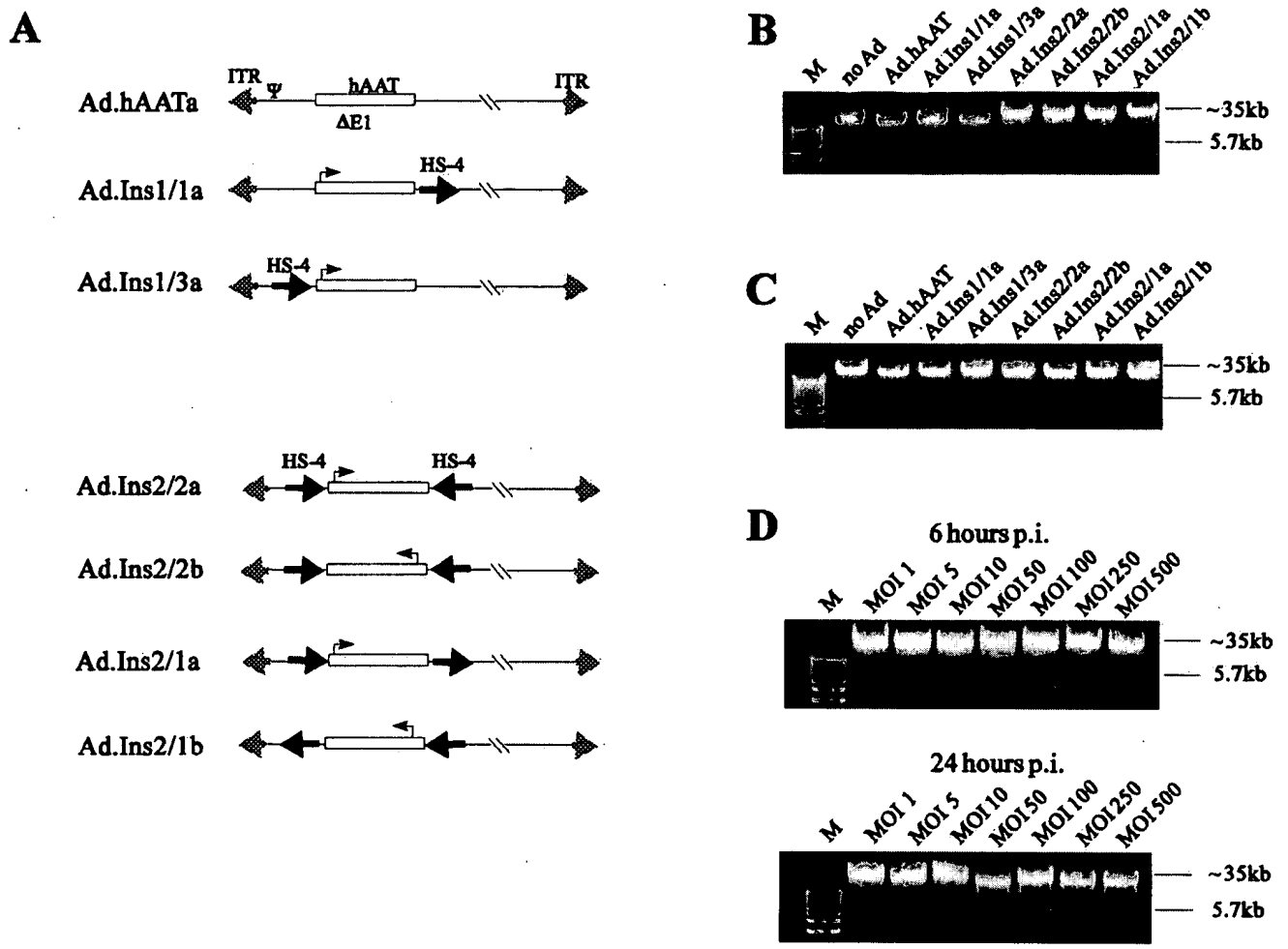


FIG. 1. Formation of genomic derivatives during replication of first-generation vectors carrying two IRs. (A) Structure of first-generation Ad vectors with transgene cassettes inserted into the E1 region. The transgene cassette (hAAT) comprises an inducible MRE promoter (33) linked to hAAT cDNA and a polyadenylation signal derived from the bPA. The IR sequence represents the 1.2-kb fragment of the HS-4 domain derived from the chicken β-globin locus (3). Ad.Ins1/1a and Ad.Ins1/3a contain only one HS-4 element upstream or downstream of the transgene cassette. Ad.Ins2/2a, Ad.Ins2/2b, Ad/Ins2/1a, and Ad.2/1b each contain two HS-4 elements as IRs or DRs, respectively. Ad.hAATa contains only the hAAT cDNA and a polyadenylation signal derived from the bPA. (B) Replication studies. 293 Cells were infected with Ads at an MOI of 10 PFU/cell. Thirty-six hours postinfection (p.i.), total DNA was isolated from infected cells and analyzed (undigested) by electrophoresis in agarose gels stained with ethidium bromide. The sample volume loaded per lane corresponds to the amount of DNA isolated from ~20,000 infected cells. The full-length (35-kb) viral genome comigrates with fragments of cellular DNA. The specific replication derivatives from Ad.Ins2/2a and -b with lengths of 5.7 kb are marked. (C) 293 cells were infected, and DNA was analyzed as in panel B. To inhibit viral replication, hydroxyurea at a final concentration of 10 mM was added to the 293 culture medium 1 h postinfection. (D) 293 cells were infected with Ad.Ins2/2a at different MOIs (PFU per cell). DNA analysis was performed as described in panel B before Ad replication started (6 h postinfection) and after the initiation of replication (24 h postinfection) (34). M, 1-kb DNA ladder (Gibco BRL, Grand Island, N.Y.)

after infection with Ad.Ins2/2a and Ad.Ins2/2b, both of which contained HS-4 elements as IRs. These bands were absent in cells infected with the control vectors or the vectors containing two HS-4 DRs. The 5.7-kb bands were identified as Ad vector derivatives by Southern blotting with a transgene specific probe (data not shown). These derivative genomes are hereafter referred to as ΔAd.IR-1 (for the Ad.Ins2/2a derivative) and ΔAd.IR-2 (for the Ad.Ins2/2b derivative).

Quantitative Southern analysis revealed that $\sim 5 \times 10^4$ of the 5.7-kb ΔAd.IR-1 or ΔAd.IR-2 genomes and $\sim 10^5$ corresponding full-length genomes were produced per cell after infection with the corresponding first-generation vector at an MOI of 10. The appearance of these small vector genomes was linked to adenoviral DNA replication, because it was absent when hydroxyurea, an inhibitor of viral DNA replication, was added to the 293 culture medium after infection (Fig. 1C). The amount of ΔAd.IR-1 genomes produced was analyzed 6 and 24 h after infection of 293 cells with different MOIs of Ad.Ins2/2a (Fig. 1D). ΔAd.IR-1 genomes were absent when cell lysates were analyzed before replication started (6 h postinfection). At 24 h postinfection, the number of ΔAd.IR-1 genomes increased when the viral dose was between MOI 1 and 10 and reached a plateau after infection with higher MOIs (50 to 500).

DNA restriction analysis and sequencing of the 5.7-kb viral genomes revealed the genome structure shown in Fig. 2A. Notably, the 4.0-kb *Not*I and the 1.4-kb *Bam*HI fragments were specific for ΔAd.IR-1 and ΔAd.IR-2 and were absent from the full-length genome and the original shuttle plasmid (Fig. 2B). Both deleted vectors, ΔAd.IR-1 and ΔAd.IR-2, contained the transgene cassette flanked by the inverted HS-4 elements, which are linked on both sides to two identical inverted copies of Ad DNA comprising the Ad ITR and packaging signal. Importantly, these genomes were devoid of all Ad sequences that encode viral proteins. This structure was confirmed by sequencing the *Not*I fragments of both ΔAd.IR genomes with

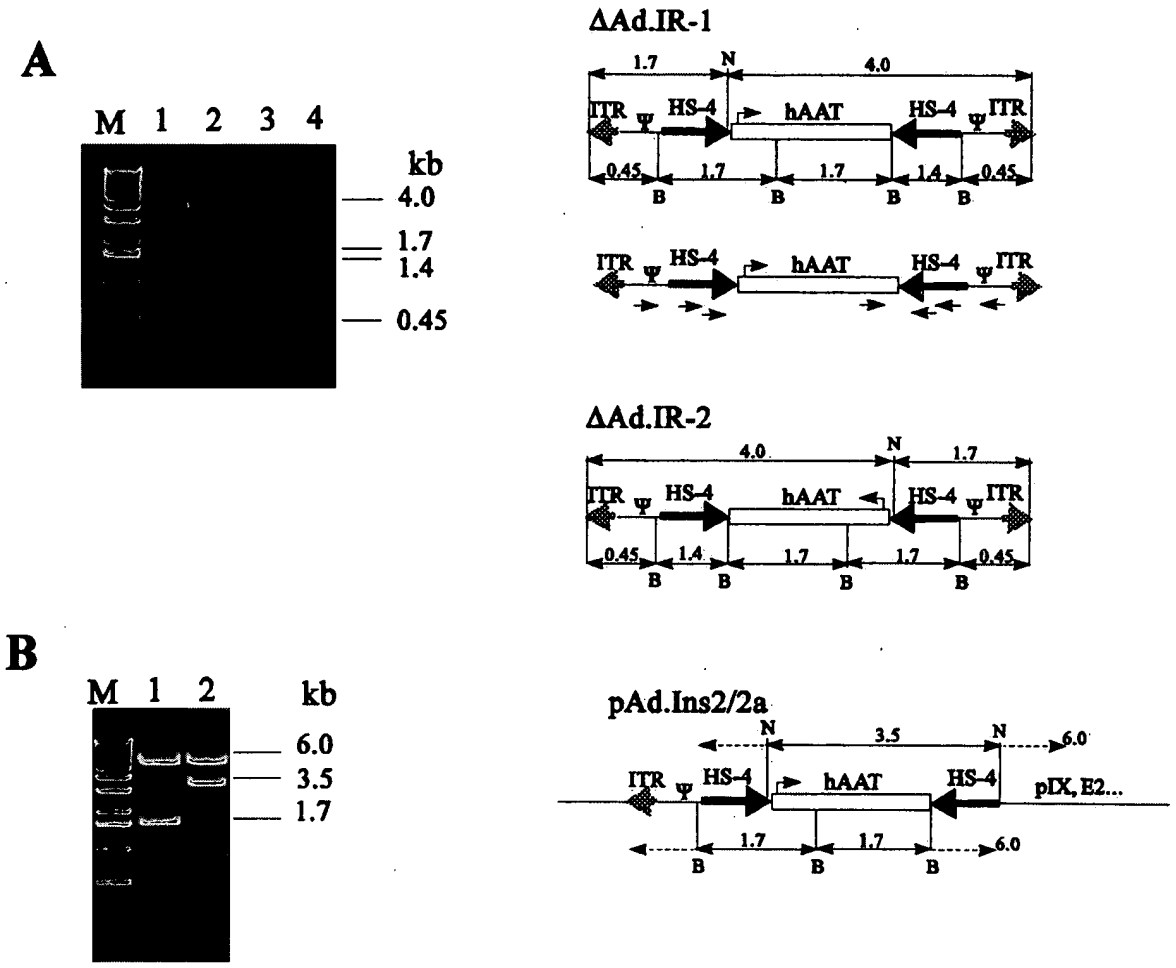


FIG. 2. Structure of ΔAd.IR-1 and ΔAd.IR-2 genomes. (A) ΔAd.IR-1 and Ad.IR-2 genomes were excised from gels after gel electrophoresis, were purified, and were submitted to restriction analysis or sequencing with primers specific to Ad sequences or to the transgene cassette (see Materials and Methods). The results of restriction analysis of ΔAd.IR-1 (lanes 1 and 2) and ΔAd.IR-2 (lanes 3 and 4) genomes after digestion with *Not*I (N) (lanes 1 and 3) or *Bam*HI (B) (lanes 2 and 4) are shown. The structures of ΔAd.IR-1 and ΔAd.IR-2 genomes were deduced from restriction and sequencing data. The localization of sequencing primers specific to nt 319 to 338 within the Ad region is indicated in the lower chart for ΔAd.IR-1. (B) For comparison, the *Bam*HI (lane 1) or *Not*I (lane 2) digest of the corresponding pΔE1sp1a-based shuttle plasmid used for generation of Ad.Ins2/2a is shown. The 3.6-kb *Not*I and 1.7-kb *Bam*HI fragments are specific for the full-length viral Ad.Ins2/2a genome. Due to additional restriction sites within the viral genome, the restriction pattern is only shown for the shuttle plasmid. The 6.0-kb fragment contains the pΔE1sp1a backbone. Ψ, Ad packaging signal; IR, HS-4 IR; hAAT, transgene expression cassette; pIX, gene for Ad pIX protein. All fragment sizes are in kilobases. M, 1-kb DNA ladder (Gibco BRL).

primers specific to the Ad packaging region or primers specific to regions within the transgene cassette (Fig. 2A). The sequencing data demonstrated an accurate mechanism for the duplication of the IRs in conjunction with the Ad packaging signal and ITR.

The appearance of ΔAd.IR genomes with identical, duplicated regions was linked to viral DNA replication and required the presence of IRs. DRs did not mediate this process. Based on these results, we hypothesized that the unique structure of ΔAd.IR could be the result of homologous recombination processes stimulated by the IRs flanking the transgene cassette (Fig. 3). This process could involve the formation of a Holliday structure, which can be resolved by a classical isomerization process (11, 16) or during Ad replication.

If this model is correct, recombination products should appear in cells coinfected with two Ads; one containing a sequence in leftward orientation, and the other containing an identical sequence in rightward orientation with respect to the Ad ITR and packaging signal (Fig. 4A). To test this hypothesis, vectors Ad.Ins1/3a and Ad.Ins1/3b containing one HS-4 element and the hAAT transgene cassette in leftward or rightward orientation, respectively, were added onto 293 cells separately or in combination. Viral DNA was analyzed together with cellular DNA after development of CPE, as described in Fig. 1. As expected, no small vector derivatives were detected in cells infected separately with Ad.Ins1/3a or Ad.Ins1/3b. Importantly, a 4.2-kb deleted vector genome was generated in cells after simultaneous infection with the two vectors (Ad.Ins1/3a plus Ad.Ins1/3b; Fig. 4B). This product can only form when the two double-stranded genomes (Ad.Ins1/3a and Ad.Ins1/3b) recombine via the HS-4-hAAT homology region. The amount of 4.2-kb deleted vector genomes was similar to the amount observed for ΔAd.IR-1/2 (Fig. 1A). A corresponding recombination product appeared in cells coinfected with Ad.Ins1/1 and Ad.Ins1/1b containing the hAAT cassette followed by the HS-4 element. To demonstrate that recombination is not associated with the specific sequence or structure of the HS-4 element and that recombination can be mediated by other sequences, vectors Ad.hAATa and Ad.hAATb were employed. These vectors contained 1.4-kb hAAT cDNA seg-

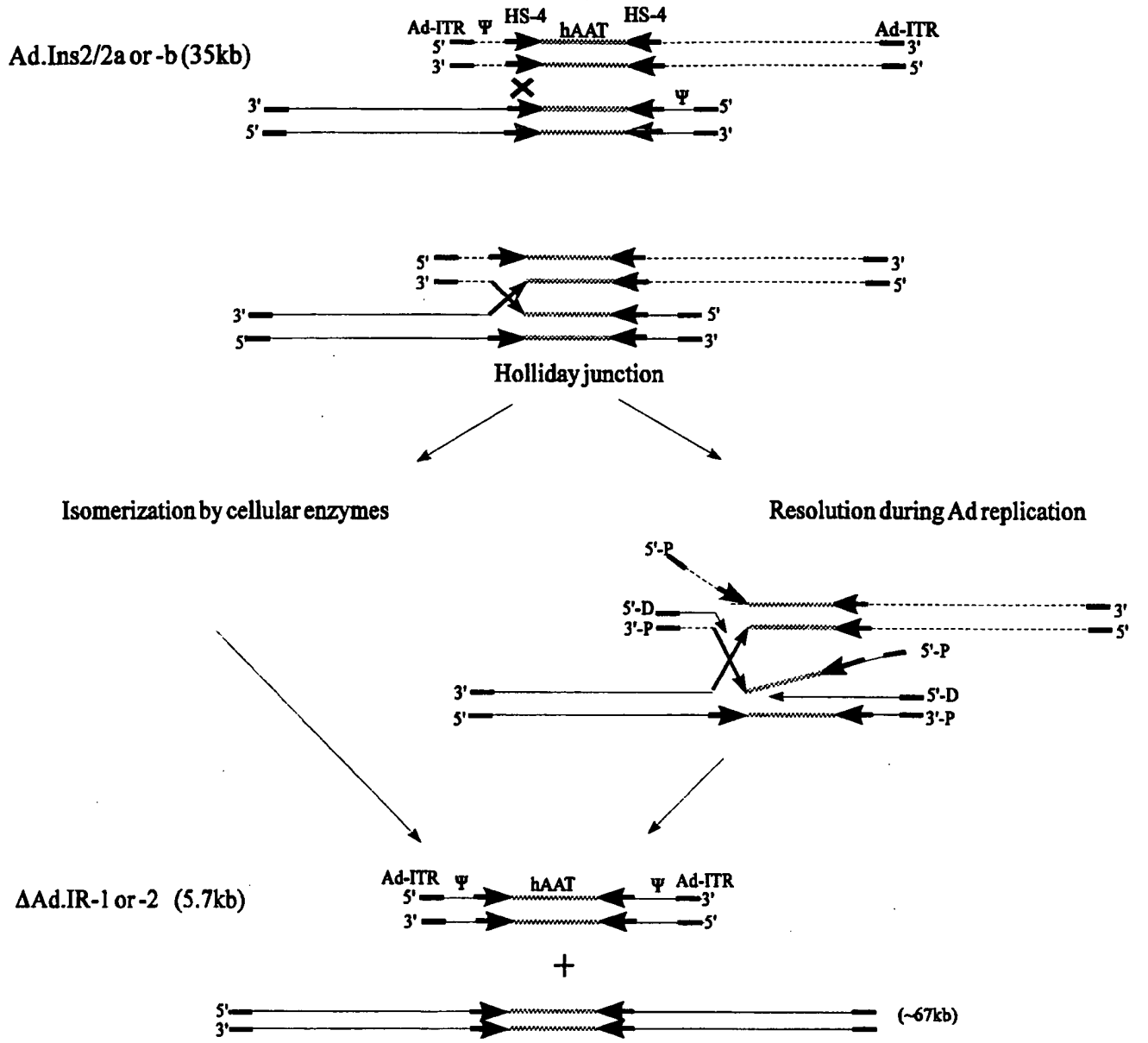


FIG. 3. Hypothetical mechanism for the formation of ΔAd.IR genomes. According to the accepted model for general recombination, the recombination process could begin with the pairing of homologous regions within the HS-4 elements of two double-stranded Ad genomes containing IRs (e.g., Ad.Ins2/2a). Then, cellular recombination enzymes would mediate the exchange of a single strand between the two double-stranded viral genomes, resulting in a Holliday junction. This structure could isomerize by undergoing a series of rotations, as described for the classical Holliday resolution model (11, 16). Alternatively, the Holliday structure could be resolved during Ad replication. The synthesis of a daughter strand (5'-D) is associated with the displacement of one parental strand (5'-P) and could occur along the crossed strands. When two oppositely moving displacement forks meet at the crossover, the two crossed parental strands will no longer be held together and will separate, resulting in partially duplex and partially single-stranded molecules. The synthesis is then completed on the displaced parental strand. A similar dissociation mechanism is described for Ad2 replication (21). One of the recombination products has the structure of ΔAd.IR genomes. Theoretically, a second double-stranded product, with a length of ~67 kb, should also form. We were not able to demonstrate the presence of this product by Southern blotting. This product is apparently not efficiently generated. Its large size would require extremely long periods of time for replication (>40 min). Based on the predicted structure, the 67-kb product would lack packaging signals. Furthermore, the large size of this product would prevent packaging. The figure shows the pairing and the cross-exchange for only one HS-4 pair. With the same likelihood, recombination can occur between the other HS-4 pair, resulting in identical products. To simplify the Holliday resolution model by Ad replication, DNA synthesis initiated from only one genome end is shown.

ments linked to 0.3-kb bPA polyadenylation signals in left- or rightward orientation. In cells coinfected with Ad.hAATa and Ad.hAATb, this 1.7-kb hAAT-bPA region of homology also efficiently mediated the formation of small deleted vectors. The structure of the deleted genomes formed after coinfection with two viruses was confirmed by restriction analysis. Representative data were shown in Fig. 4B.

In conclusion, vectors deleted for all viral genes are efficiently formed by a process that appears to involve homologous recombination between two IRs present in one vector or by recombination between independent vectors, each containing one inverse homology element. Importantly, in contrast to recombination within a single parental vector carrying two IRs, recombination between two coinfected vectors results in the formation of a hybrid ΔAd.IR carrying elements from both parental vectors.

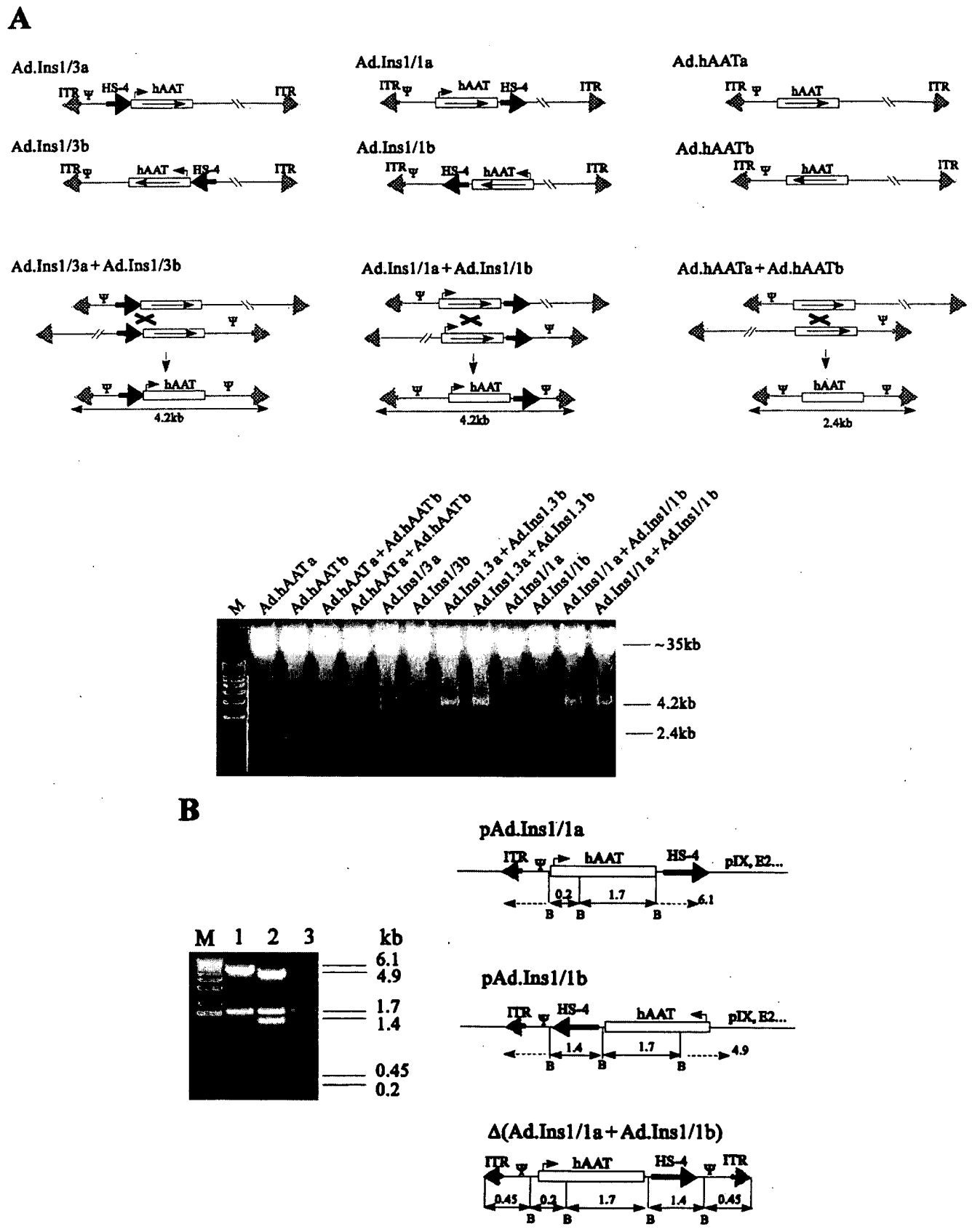


FIG. 4. Formation of genomic derivatives by coinfection of two parental vectors each carrying only one homology element. (A) The structures of Ad.hAATa, Ad.Ins1/1a, and Ad.Ins1/3a are described in Fig. 1A. Ad.hAATb, Ad.Ins1/1b, and Ad.Ins1/3b contain their individual inserts in opposite orientations with respect to their counterparts Ad.hAATa, Ad.Ins1/1a, and Ad.Ins1/3a. Coinfection of Ad.hAATa/b, Ad.Ins1/1a/b, or Ad.Ins1/3a/b allows for intermolecular recombination and

The recombination model for the formation of the described ΔAd.IR genomes relies on an extended homology between the IRs. In order to assess whether the formation of genomic derivatives quantitatively depends on the length of the homologous elements, additional vectors were generated that contained shorter IR elements than Ad.Ins2/2a (Fig. 5). The shorter IRs derived from the HS-4 fragments, with lengths of 1.0 or 0.5 kb, allowed for the generation of the ΔAd.IR genomes at the same rate as the 1.2-kb HS-4 elements used in the experiment described in Fig. 1A. The corresponding replication derivatives had the expected sizes of 3.9 and 4.9 kb, respectively. However, a vector that contained very short IRs consisting of stretches with 20 bp of dC and dG did not yield detectable replication derivatives, indicating that certain lengths of IRs are required for ΔAd.IR formation. Notably, both the 1.0- and 0.5-kb-long IRs were deleted for a GC-rich region located at the terminal ends of the 1.2-kb HS-4 fragments (3) (Fig. 5). Together with the data shown in Fig. 4, this suggests that formation of ΔAd.IR genomes can be mediated by any sequence present as an IR in the E1 region of Ad vectors. To test whether the sequence of the intervening region between the IRs is critical for the formation of genomic derivatives, a vector was produced that had two inverted 1.2-kb HS-4 fragments flanking a 4.1-kb β-Gal gene. During replication of this vector, the expected 8.2-kb genome was formed as efficiently as the genome of Ad.IR-1 or Ad.IR-2. This demonstrates that IRs can be employed for the generation of deleted vectors containing a transgene cassette of choice.

The presence of packaging signals in ΔAd.IR-1 and ΔAd.IR-2 prompted us to test whether these viral genomes were packaged into virions that could be banded in CsCl gradients. 293 cells infected with Ad.Ins2/2a (MOI 10) were lysed 48 h postinfection in order to liberate produced viral particles, and cell lysates were separated by ultracentrifugation in CsCl gradients to band and visualize viral particles. A prominent additional viral band with a buoyant density of ~1.32 g/cm^3 appeared in CsCl step gradients between virions containing full-length (~35-kb) genomes and empty or defective particles (Fig. 6A). Analysis of viral material from purified particles contained in this band demonstrated the packaged 5.7-kb ΔAd.IR-1 genome (Fig. 6B). ΔAd.IR-1 particles could be separated from contaminating particles with packaged full-length genomes (Ad.Ins2/2a) by an additional CsCl equilibrium gradient (Fig. 6B, lane 3). Based on quantitative Southern analysis of DNA isolated from CsCl banded particles, ~10^4 packaged ΔAd.IR-1 genomes and ~5 × 10^4 packaged full-length genomes were produced per cell (data not shown). Considering the amount of corresponding genomes found at the time of virus harvest inside the cell (Fig. 1B), this indicates that both the 5.7- and the 35-kb genomes were packaged efficiently. Final preparations of ΔAd.IR-1 after two CsCl gradient purifications contained less than 0.1% contaminating first-generation Ad.Ins2/2a, as analyzed by plaque assay (see Materials and Methods). CsCl purification of other ΔAd.IR vectors resulted in similar findings (data not shown).

Electron microscopy of ΔAd.IR-1 particles demonstrated the same icosahedral shape as the first-generation, Ad.Ins2/2a, particles (Fig. 7). Staining with uranyl acetate allows the central viral cores to appear electron dense. While the lumina of particles containing the full-length genomes were homogenously electron dense, virions containing the smaller genomes had only spotted luminal staining, indicating the presence of less packaged DNA in ΔAd.IR-1 particles. We speculate that only one deleted genome is packaged per virion.

As a further test for the intactness of ΔAd.IR-1 particles, we measured the ability to mediate gene transfer into cultured cells based on reporter gene (hAAT) expression (Fig. 8). Confluent SKHEp-1 cells were infected with purified ΔAd.IR-1 and Ad.Ins2/2a particles at an MOI of 2,000 genomes per cell. This cell line does not significantly support the replication of first-generation Ad (30). The level of hAAT expression at day 3 after infection was comparable for both vectors, indicating that in vitro gene transfer was similarly efficient. While transgene expression from the full-length vector was stable during the analyzed time period (7 days), hAAT expression declined gradually for ΔAd.IR-1 starting at day 4 postinfection. Southern analysis of viral DNA isolated from infected cells revealed that the short duration of transgene expression was due to the instability of ΔAd.IR-1 genomes within transduced cells (Fig. 9). The concentration of full-length viral genomes (Ad.Ins2/2a) was comparable in cells harvested at day 1 and day 7 postinfection. In contrast, while the input concentration of ΔAd.IR-1 genomes analyzed at day 1 postinfection was as high as for first-generation vectors, the number of ΔAd.IR-1 vector genomes was barely detectable in transduced cells at day 7 postinfection.

In conclusion, small ΔAd.IR genomes are efficiently formed and packaged during replication of E1-deleted Ad vectors containing two IRs flanking a reporter gene cassette. The mechanism of formation of ΔAd.IR requires viral DNA replication and most likely involves homologous recombination. ΔAd.IR formation can be achieved using IRs of various lengths and origins. The inverted homology elements required for ΔAd.IR generation can also be provided in *trans* by the coinfection of two independent viruses, resulting in a hybrid ΔAd.IR. Particles containing the small genomes devoid of all viral genes could be separated from virions with full-length genomes based on their lighter buoyant density. These particles infected cultured cells with the same efficiency as first-generation vectors; however, deleted genomes were only short-lived within transduced cells. The production of high titers by using two IRs is technically straightforward and does not require helper viruses, because all functions required for the replication of the small genome and for particle formation are provided from the full-length genomes amplified in the same cell.

DISCUSSION

We demonstrated that IRs inserted into first-generation Ad vector genomes mediated precise genomic rearrangement, resulting in vector genomes that were devoid of all viral genes and which were efficiently packaged into functional Ad capsids. This finding has practical implications for Ad vector develop-

production of the indicated products, since the inverse homology elements of a/b vector combinations can pair as proposed for vectors carrying IRs (Fig. 3). 293 cells were infected with the indicated vector combinations at an MOI of 50, and viral DNA was analyzed as described in Fig. 1. Infection with single vector types carrying one homology element did not yield any genomic derivatives. Infection with a/b vector pairs generated the expected vector derivatives of 2.4 kb (Ad.hAATa/b) and 4.2 kb (Ad.Ins1/1a/b and Ad.Ins1/3a/b). (B) The structure of the subgenomic (4.2-kb) Ad genome resulting from coinfection of Ad.Ins1/1a and Ad.Ins1/1b was analyzed by *Bam*HI restriction analysis as described in Fig. 2. For comparison, the *Bam*HI digests of the corresponding pΔE1sp1a-based shuttle plasmids used for the generation of Ad.Ins1/1a (pAd.Ins1/1a, lane 1) and Ad.Ins1/1b (pAd.Ins1/1b, lane 2) are shown. The *Bam*HI fragments specific for the 4.2-kb recombination product [Δ(Ad.Ins1/1a + Ad.Ins1/1b)] (lane 3) represent the double 0.45-kb band and the combination of the 0.2-, 1.4-, and 1.7-kb bands.

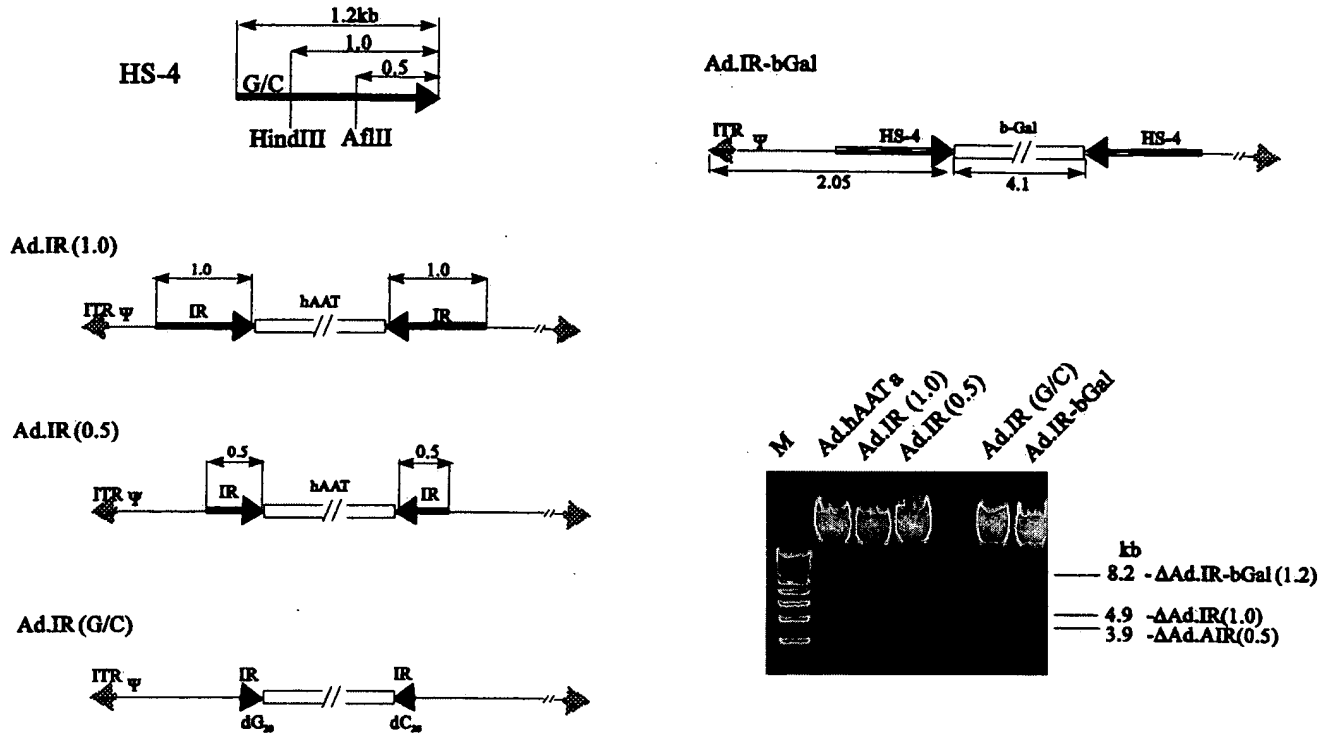


FIG. 5. Effect of shortened IRs on the formation of vector derivatives. 293 cells were infected with Ad vectors, and the viral DNA was analyzed as described in Fig. 1. Ad.IR(1.0) contained HS-4-derived IRs with lengths of 1 kb. Ad.IR(0.5) contained 0.5-kb-long IRs. The restriction sites for *Hin*dIII and *Afl*II present within the 1.2-kb HS-4 element that were used to produce the shorter IRs are indicated. The proximal 200 nt of the HS-4 fragment contain a GC-rich region (chart at the top of the figure). This region and about 150 bp of flanking polylinker are deleted in Ad.IR(1.0) and Ad.IR(0.5). The vector Ad.IR(G/C) contains synthetic 20-mer dG and dC stretches flanking the transgene cassette. The vector derivatives have lengths of 4.9 kb for Ad.IR-1(1.0) and 3.9 kb for Ad.IR(0.5). The structure of the parental vectors is shown on the right. Ad.bGal contained the 1.2-kb HS-4 elements flanking a 4.1-kb *lacZ* cassette inserted into pAd.RSV (14). The corresponding rearranged vector derivative had the expected length of 8.2 kb.

ment and may contribute to a better understanding of Ad replication and the functional importance of IRs.

ΔAd.IR genomes contained only the transgene cassette flanked on both sides by duplicated IRs, Ad packaging signals, and ITRs. This specific structure could be generated precisely and reproducibly by using IRs of different sizes and origins, but not by DRs. These findings implicated the involvement of homologous recombination in the formation process (Fig. 3).

This hypothesis was confirmed by coinfection with two vectors, each containing only one region of homology. The ΔAd.IR genomes detected after coinfection could only have formed by both vectors recombining through oppositely orientated homology elements. This genomic rearrangement process could involve the formation of a Holliday structure whose formation and stabilization could be supported by the Ad DBP, which is known to enhance intermolecular interactions (39). We postulate that the Holliday structure could be resolved by classical isomerization mediated by cellular recombination enzymes, which are highly conserved during evolution (11, 16). Alternatively, the unique mechanism of Ad DNA synthesis may account for the efficient resolution of a Holliday structure, as outlined in Fig. 3. In this context, it would be interesting to test whether similar genomic rearrangements mediated by IRs can be achieved with other DNA viruses (HSV, SV40, or polyomavirus) which use different replication strategies.

At the late stages of viral infection with a relatively low MOI, $\sim 5 \times 10^4$ ΔAd.IR genomes were produced per cell, which is only twofold less than the number of full-length genomes produced per cell. This implies that either the event that forms ΔAd.IR genomes occurs very frequently or that only a small number of rearranged genomes are originally formed and later amplified by the Ad replication machinery. The ΔAd.IR genomes are approximately six times shorter than the full-length genomes and could therefore have a replicative advantage. Previous studies have demonstrated that small Ad vector genomes are replicated by viral proteins expressed from full-length genomes present within the same infected cell (4, 6, 9, 26, 27). This supports the hypothesis that the vector rearrangement is a rare event and that ΔAd.IR genomes are amplified together with full-length genomes in transduced cells. This is further supported by the low frequency of recombination seen between Ad shuttle plasmids used for the generation of recombinant Ads. The critical importance of ΔAd.IR genome replication is also underscored by the observation that the amount of generated ΔAd.IR genomes correlated with the kinetics of Ad replication. The number of ΔAd.IR genomes generated increased with the viral dose between 1 and 10 PFU/cell and reached a plateau when infection MOIs were greater than 10. In this context, it is notable that Ad replication starts only if a certain threshold of early viral protein has accumulated and reaches a plateau that is dictated by limiting viral and cellular factors (37).

Our data does not exclude other mechanisms for the formation of ΔAd.IR genomes. Particularly intriguing is the unique mechanism of Ad replication, involving single-stranded intermediates, that can form intramolecular secondary structures. In this context, stem-loop or cruciform-like structures formed through intrastrand hybridization of IRs may be functionally important in the formation of ΔAd.IR genomes. Elongation by

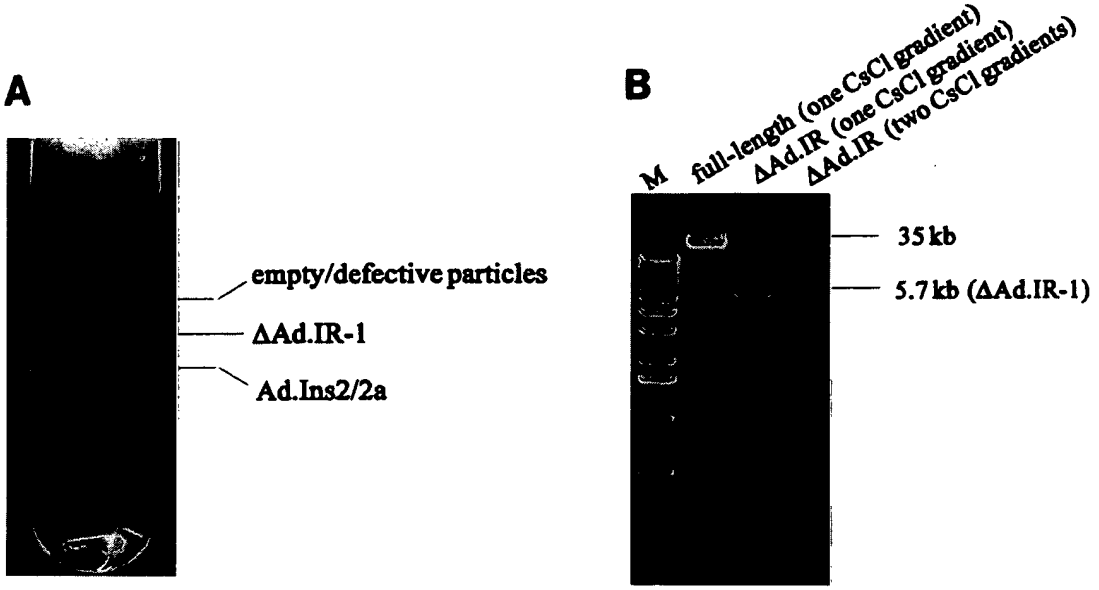


FIG. 6. Isolation of particles containing packaged ΔAd.IR-1 genomes. (A) Ad.Ins2/2a was amplified in 293 cells and banded by ultracentrifugation in a one-step CsCl gradient. The lower band contains full-length genome and Ad.Ins2/2a particles, the middle band contains ΔAd.IR-1 particles, and the top band contains empty and defective viral particles. Viral material from 3×10^8 293 cells collected 36 h after infection with Ad.Ins2/2a at an MOI of 50 PFU/cell is shown. (B) Undigested viral DNA purified from banded viral material after centrifugation in CsCl gradients was analyzed in 0.8% agarose gels. Lane 1, DNA isolated from 10 μl of banded Ad.Ins2/2a (full-length genome, 35 kb); lanes 2 and 3, DNA isolated from 10 μl of the banded ΔAd.IR-1 particles (5.7-kb genome) after a one-step CsCl gradient (lane 2) or after additional purification of ΔAd.IR-1 virions by ultracentrifugation in a CsCl equilibrium gradient (lane 3).

Ad Pol is relatively slow, and it may provide a lag time sufficient to form stem structures within single strands during their displacement. Hypothetically, Ad Pol can pause at the IR-stimulated hairpin structure and switch template strands. Similar mechanisms have been described for other DNA polymerases (1, 13, 19, 38) and for retroviral reverse transcriptases (7, 12, 15). Although the involvement of intramolecular stem-loop or cruciform-like structures formed by IRs present within single-stranded replication intermediates appeared to be an attractive basis for the explanation of the ΔAd.IR structure, the data obtained with coinfected Ad viruses containing only single homology regions contradicted this hypothesis. Nonetheless, local formation of intrastrand secondary structures may initiate or support recombination processes. Clearly, our data demonstrates that Ad replication is required for the high-level production of ΔAd.IR genomes in infected cells, either as the etiological event responsible for the genomic rearrangements or as a supportive mechanism for the amplification of rearranged genomes.

The structure of ΔAd.IR particles revealed by electron microscopy and their buoyant density in CsCl gradients clearly differ from empty particles. Furthermore, we demonstrated that DNase I-treated ΔAd.IR virions efficiently transferred their genomes into cells, as shown by Southern blotting and transgene expression. These facts prove that ΔAd.IR vector genomes, which contain two Ad packaging signals, were packaged into Ad capsids. While the number of deleted ΔAd.IR and corresponding full-length genomes produced per cell differed only by a factor of 2, the ratio of full-length genome particles to ΔAd.IR particles in CsCl gradients was 5:1 to 10:1. This indicates that packaging of the small genomes was 2.5- to 5-fold less efficient than that of full-length genomes. These numbers are in agreement with a study by Parks and Graham (31) in which plasmids carrying Ad genomes of different sizes were used in combination with helper virus to determine the lower packaging limit for Ad vectors. Vectors of fewer than 27 kb were recovered with about half the efficiency of larger vectors. Interestingly, a 15-kb genome was packaged at a higher efficiency than were the 20- to 25-kb-long vectors. However, the work of Parks and Graham demonstrated a clear disadvantage in the amplification of genomes of less than 25 kb during multiple virus passages. Yet, from this experiment, it was not clear whether the smaller vectors were less efficiently replicated or less efficiently packaged. The results of that study are difficult to compare with those of our ΔAd.IR vectors, which start out full length and are deleted in the producer cells (perhaps after a critical event required for packaging has occurred) during one round of large-scale amplification. Packaging of a 9-kb mini-Ad vector generated by Cre-lox recombination (27) or of encapsidated Ad chromosomes (4, 6, 17) has been previously reported.

ΔAd.IR vectors were produced by standard techniques for first-generation adenovirus amplification and purification. All

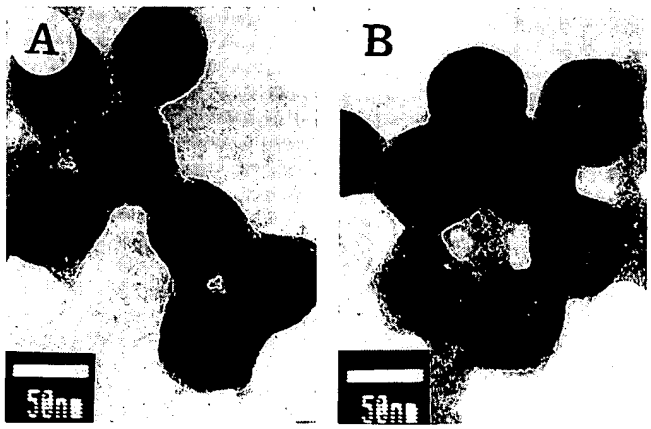


FIG. 7. Electron microscopy of Ad particles. Particles with Ad.Ins2/2a (A) and ΔAd.IR-1 (B) genomes. Virions were purified by two rounds of ultracentrifugation CsCl gradients. Magnification, ×100,000.

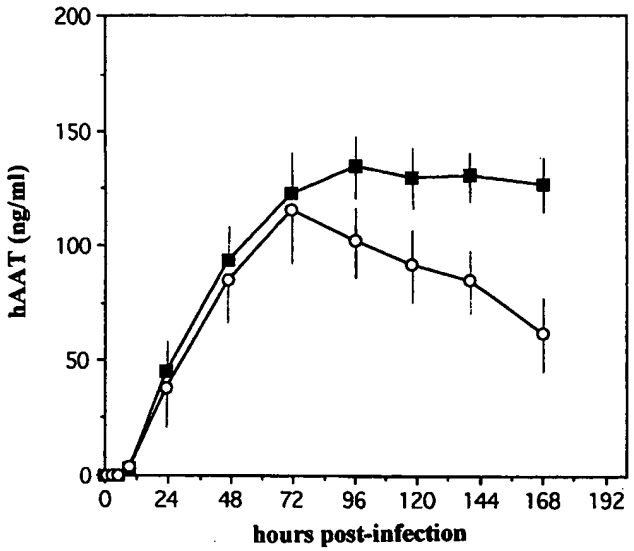


FIG. 8. Expression of hAAT after infection with Ad.Ins2/2a and ΔAd.IR-1. Confluent SKHEP-1 cells (10^6) were infected with purified Ad.Ins2/2a and ΔAd.IR-1 virus at an MOI of 2,000 genomes per cell, and hAAT concentrations were determined in culture supernatants by ELISA. Culture media was supplemented with 150 μM $ZnSO_4$ for induction of hAAT expression and was changed daily. Filled squares, cells infected with Ad.Ins2/2a; empty circles, cells infected with ΔAd.IR-1. Results are based upon three independent experiments.

the functions required for ΔAd.IR genome generation and replication and particle formation are provided from the full-length genomes amplified in the same cell. The efficiency of vector production was 10^4 packaged genomes per cell or $>1 \times 10^{13}$ packaged genomes produced in a large-scale preparation after one round of infection with first-generation vector. Banded particles containing the genomic derivatives were clearly separated in CsCl gradients based on their lighter buoyant density, which allowed for their purification without contamination with first-generation virus containing full-length genomes (Fig. 2B). In this context, the production of high-titer vectors devoid of all viral genes is less labor intensive than helper-dependent production of gutless vectors (32) or packaged adenovirus minichromosomes (17), both of which require multiple passages of virus.

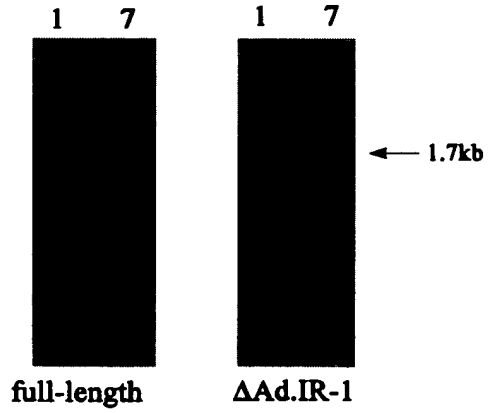


FIG. 9. Analysis of viral DNA in transduced cells. Confluent SKHEP-1 cells were infected with purified Ad.Ins2/2a and ΔAd.IR-1 virus at an MOI of 2,000 genomes per cell. At days 1 and 7 after infection (indicated on top), genomic DNA was extracted. Ten-microgram samples were digested with *Bam*HI and analyzed by Southern blotting with an hAAT-specific probe. The vector-specific fragment is 1.7 kb.

The ΔAd.IR-1 vector infected cells with the same efficiency as the corresponding first-generation vector based on a similar level of reporter gene expression at day 3 postinfection. However, transgene expression declined over time due to the instability of ΔAd.IR-1 genomes in transduced cells. Nonetheless, the high infectivity of ΔAd.IR-1 indicates that the viral structural elements present in ΔAd.IR particles are functionally intact and mediate efficient cell entry, endosomal lysis, and nuclear import. This may allow for the efficient infection of a variety of cell types, including nondividing cells. The potential for highly efficient gene transfer, together with the fact that the ΔAd.IR vector genomes lack viral genes, make ΔAd.IR vectors practically important. For example, a transient transgene expression would be sufficient for a variety of cell biology or cell cycle studies which require efficient gene transfer and the absence of side effects associated with the expression of adenoviral proteins.

On the other hand, approaches aimed toward gene therapy of genetic disorders require stable gene expression. In agreement with the data presented here, we recently demonstrated that a 9-kb mini-Ad genome generated by Cre-lox recombination was packaged into Ad particles that transduced cells efficiently; however, they were short lived and were completely degraded by day 7 after in vitro infection (27). We suggested that the expression of certain viral proteins, including pTP, is required to confer genome stability in transduced cells (26). Furthermore, in a study related to the present paper, we utilized adeno-associated virus elements in combination with the described rearranged vectors to mediate integration as a means of vector stabilization allowing for stable transgene expression.

The strategy of ΔAd.IR generation using coinfection of viruses each containing one inverse homology element can be used to combine elements of choice from two independent viruses. Because the packaging capacity of both parental vectors could be exploited, large inserts could be accommodated by ΔAd.IR vectors. In order to test this, we are currently generating ΔAd.IR vectors containing 12- to 15-kb transgenes. Furthermore, promoter and transgene could be placed into separate vectors so that the transgene would not be expressed during generation and amplification of the parental vectors unless both vectors were coinfected. This strategy may be extremely useful whenever transgene expression is toxic to producer cells.

This study demonstrates proof of the principle that IRs can be used to create predictable genetic rearrangements within the framework of Ad replication. This method allows for the reliable and efficient generation of vectors devoid of all viral genes and has potential application in the development of vectors for gene therapy.

ACKNOWLEDGMENTS

We thank Zong-Yi Li for technical assistance. We are grateful to Dmitry Shayakhmetov and David Russell for critical discussion.

This work was supported by the Cystic Fibrosis Foundation and NIH grants R01 CA80192-01 and R21 DK55590-01. D.S.S. is a recipient of a predoctoral DAAD fellowship. C.A.C. is supported by a grant from the University of Washington (Department of Environmental Pathology and Toxicology).

REFERENCES

1. **Bedinger, P., M. Munn, and B. M. Alberts.** 1989. Sequence-specific pausing during in vitro DNA replication on double-stranded DNA templates. J. Biol. Chem. 264:16880–16886.
2. **Challberg, M. D., and T. J. Kelly.** 1989. Animal virus DNA replication. Annu. Rev. Biochem. 58:671–717.
3. **Chung, J. H., A. C. Bell, and G. Felsenfeld.** 1997. Characterization of the

chicken beta-globin insulator. Proc. Natl. Acad. Sci. USA 94:575–580.
4. **Fisher, K. J., H. Choi, J. Burda, S. Chen, and J. M. Wilson.** 1996. Recombinant adenovirus deleted of all genes for gene therapy of cystic fibrosis. Virology 217:11–22.
5. **Gordenin, D. A., K. S. Lobachev, N. P. Degtyareva, A. L. Malkova, E. Perkins, and M. A. Resnick.** 1993. Inverted DNA repeats: a source of eukaryotic genomic instability. Mol. Cell. Biol. 13:5315–5322.
6. **Haecker, S. E., H. H. Stedman, R. J. Balice-Gordon, D. B. J. Smith, J. P. Greelish, M. A. Mitchell, A. Wells, H. L. Sweeney, and J. M. Wilson.** 1996. In vivo expression of full-length human dystrophin from adenoviral vectors deleted for all viral genes. Hum. Gene Ther. 7:1907–1914.
7. **Harrison, G. P., M. S. Mayo, E. Hunter, and A. M. L. Lever.** 1998. Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5′ and 3′ of the catalytic site. Nucleic Acids Res. 26:3433–3442.
8. **Hauser, M. A., A. Amalfitano, R. Kumar-Singh, S. D. Hauschka, and J. S. Chamberlain.** 1997. Improved adenoviral vectors for gene therapy of Duchenne muscular dystrophy. Neuromusc. Disord. 7:277–283.
9. **Hay, R. T., N. D. Stow, and I. M. McDougall.** 1984. Replication of adenovirus mini-chromosomes. J. Mol. Biol. 175:493–510.
10. **Heffelfinger, S. C., H. H. Hawkins, J. Barrish, L. Taylor, and G. J. Darlington.** 1992. SK HEP-1: a human cell line of endothelial origin. In Vitro Cell. Dev. Biol. 28A:136–142.
11. **Holliday, R.** 1990. The history of the DNA heteroduplex. Bioassays 12:133–142.
12. **Ji, J., J. S. Hoffmann, and L. Loeb.** 1994. Mutagenicity and pausing of HIV reverse transcriptase during HIV plus-strand DNA synthesis. Nucleic Acids Res. 22:47–52.
13. **Kaguni, L. S., and D. A. Clayton.** 1982. Template-directed pausing in in vitro DNA synthesis by DNA polymerase A from Drosophila melanogaster embryos. Proc. Natl. Acad. Sci. USA 79:983–987.
14. **Kay, M. A., F. Graham, F. Leland, and S. L. Woo.** 1995. Therapeutic serum concentrations of human alpha1-antitrypsin after adenoviral-mediated gene transfer into mouse hepatocytes. Hepatology 21:815–819.
15. **Kim, J. K., C. Palaniappan, W. Wu, P. J. Fay, and R. A. Bambara.** 1997. Evidence for a unique mechanism of strand transfer from the transactivation response region of HIV-1. J. Biol. Chem. 272:16769–16777.
16. **Kowalczykowski, S. C.** 1991. Biochemistry of genetic recombination: energetics and mechanism of DNA strand exchange. Annu. Rev. Biophys. Chem. 20:539–575.
17. **Kumar-Singh, R., and D. B. Farber.** 1998. Encapsidated adenovirus minichromosome-mediated delivery of genes to the retina: application to the rescue of photoreceptor degeneration. Hum. Mol. Genet. 7:1893–1900.
18. **Kumar-Singh, R., and J. S. Chamberlain.** 1996. Encapsidated adenovirus mini-chromosomes allow delivery and expression of a 14kb dystrophin cDNA to muscle cells. Hum. Mol. Genet. 5:913–921.
19. **LaDuca, R. J., P. J. Fay, C. Chuang, C. S. McHenry, and R. A. Bambara.** 1983. Site-specific pausing of DNA synthesis catalyzed by four forms of E.coli DNA polymerase III. Biochemistry 22:5177–5188.
20. **Leach, D. R.** 1994. Long DNA palindromes, cruciform structures, genetic instability, and secondary structure repair. Bioassays 16:893–900.
21. **Lechner, R. L., and T. J. Kelly, Jr.** 1977. Th structures of replicating adenovirus 2 DNA molecules. Cell 12:1007–1020.
22. **Leegwater, P. A. J., R. F. A. Romboufs, and P. C. van der Vliet.** 1988. Adenovirus DNA replication in vitro: duplication of single-stranded DNA containing a panhandle structure. Biochim. Biophys. Acta 951:403–410.
23. **Lieber, A., and M. A. Kay.** 1996. Adenovirus-mediated expression of functional ribozymes in mice. J. Virol. 70:3153–3158.
24. **Lieber, A., M. J. Vrancken Peeters, and M. Kay.** 1995. Adenovirus-mediated transfer of the amphotropic retrovirus receptor cDNA increases retroviral transduction in cultured cells. Hum. Gene Ther. 6:5–11.
25. **Lieber, A., D. S. Steinwaerder, C. A. Carlson, and M. A. Kay.** 1999. Integrating adenovirus–adeno-associated vectors devoid of all viral genes. J. Virol. 73:9314–9324.
26. **Lieber, A., C.-Y. He, and M. A. Kay.** 1997. Adenoviral preterminal protein stabilizes mini-adenoviral genomes in vitro and in vivo. Nat. Biotechnol. 15:1383–1387.
27. **Lieber, A., C.-Y. He, I. Kirillova, and M. A. Kay.** 1996. Recombinant adenoviruses with large deletions generated by Cre-mediated excision exhibit different biological properties compared with first-generation vectors in vitro and in vivo. J. Virol. 70:8944–8960.
28. **Lieber, A., C.-Y. He, L. Meuse, D. Schowalter, I. Kirillova, B. Winther, and M. A. Kay.** 1997. The role of Kupffer cell activation and viral gene expression in early liver toxicity after infusion of recombinant adenovirus vectors. J. Virol. 71:8798–8807.
29. **Lyu, Y. L., C-T. Lin, and L. F. Liu.** 1999. Inversion/dimerization of plasmids mediated by inverted repeats. J. Mol. Biol. 285:1485–1501.
30. **Nelson, J., and M. A. Kay.** 1997. Persistence of recombinant adenovirus in vivo is not dependent on vector replication. J. Virol. 71:8902–8907.
31. **Parks, R. J., and F. L. Graham.** 1997. A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. J. Virol. 71:3293–3298.
32. **Schiedner, G., N. Morral, R. J. Parks, Y. Wu, S. C. Koopmans, C. Langston, F. L. Graham, A. L. Beaudet, and S. Kochanek.** 1998. Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. Nat. Genet. 18:180–183.
33. **Searle, P. F., G. W. Stuart, and R. D. Palmiter.** 1987. Metal regulatory elements of the mouse metallothionein-I gene. EXS (Basel) 52:407–414.
34. **Shenk, T.** 1996. *Adenoviridae*, p. 2111–2148. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fields virology, vol. 2. Lippincott-Raven Publishing, Philadelphia, Pa.
35. **Steinwaerder, D. S., and A. Lieber.** Insulation from viral transcriptional regulatory elements improves inducible transgene expression from adenoviral vectors in vitro and in vivo. Submitted for publication.
36. **Stow, N. D.** 1981. The infectivity of adenovirus genomes lacking DNA sequences from their left-hand termini. Nucleic Acids Res. 10:5105–5119.
37. **van der Vliet, P. C.** 1995. Adenovirus DNA replication, p. 1–31. *In* P. B. W. Doerfler (ed.), The molecular repertoire of adenoviruses, vol. 2. Springer Verlag, Berlin, Germany.
38. **Weaver, D. T., and M. L. DePamphilis.** 1984. The role of palindromic and non-palindromic sequences in arresting DNA synthesis in vitro and in vivo. J. Mol. Biol. 180:961–986.
39. **Zijderfeld, D. C., M. H. Stuiver, and P. C. van der Vliet.** 1993. The adenovirus DNA binding protein enhances intermolecular renaturation but inhibits intramolecular renaturation. Nucleic Acids Res. 21:641–647.

MOLECULAR AND CELLULAR BIOLOGY, Sept. 1993, p. 5315–5322
0270-7306/93/095315-08$02.00/0

Vol. 13, No. 9

Inverted DNA Repeats: a Source of Eukaryotic Genomic Instability

DMITRY A. GORDENIN,[1,2]* KIRILL S. LOBACHEV,[1] NATASHA P. DEGTYAREVA,[1] ANNA L. MALKOVA,[1] EDWARD PERKINS,[2] AND MICHAEL A. RESNICK[2]

*Department of Genetics, St. Petersburg State University, St. Petersburg 199034, Russia,[1] and Laboratory of Molecular Genetics, National Institute of Environmental Health Science, P.O. Box 12233, Research Triangle Park, North Carolina 27709[2]**

Received 20 April 1993/Returned for modification 1 June 1993/Accepted 8 June 1993

While inverted DNA repeats are generally acknowledged to be an important source of genetic instability in prokaryotes, relatively little is known about their effects in eukaryotes. Using bacterial transposon Tn*5* and its derivatives, we demonstrate that long inverted repeats also cause genetic instability leading to deletion in the yeast *Saccharomyces cerevisiae*. Furthermore, they induce homologous recombination. Replication plays a major role in the deletion formation. Deletions are stimulated by a mutation in the DNA polymerase δ gene (*pol3*). The majority of deletions result from imprecise excision between small (4- to 6-bp) repeats in a polar fashion, and they often generate quasipalindrome structures that subsequently may be highly unstable. Breakpoints are clustered near the ends of the long inverted repeats (<150 bp). The repeats have both intra- and interchromosomal effects in that they also create hot spots for mitotic interchromosomal recombination. Intragenic recombination is 4 to 18 times more frequent for heteroalleles in which one of the two mutations is due to the insertion of a long inverted repeat, compared with other pairs of heteroalleles in which neither mutation has a long repeat. We propose that both deletion and recombination are the result of altered replication at the basal part of the stem formed by the inverted repeats.

Inverted repeats with various levels of homology are common in the DNA of many eukaryotes and also in human DNA and could be an important source of genetic instability within eukaryotic genomes or within fragments cloned into other organisms. Long inverted repeats (LIRs) and short inverted repeats in prokaryotes have unique genetic properties leading to precise and imprecise excision of the regions associated with the repeats. For example, the 1.5-kb LIRs of insertions of bacterial transposons Tn*5* and Tn*10* lead to precise and imprecise excision through small (4- to 10-bp) direct repeats at or near the base of the LIRs (references 3 and 22 and references therein). Inverted repeats that are either perfect palindromes or quasipalindromes (containing a short spacer between the repeats) frequently undergo deletion (10, 11).

In order to examine the genetic consequences of LIRs in eukaryotes, we examined the fate of Tn*5* inserted into the *LYS2* gene of the yeast *Saccharomyces cerevisiae* (16–18). The low level of excision was greatly enhanced in strains containing mutations in the replicative DNA polymerase genes α and δ, suggesting a role for replication. Because the system was limited to the study of precise excision events (i.e., restoration of the Lys$^+$ phenotype), the more general consequences of LIRs, such as imprecise excision, stability of altered forms of the LIRs, and interchromosomal recombination as a more global effect of LIRs, could not be addressed. A LIR that included IS*50* repeats of Tn*5* and an internal genetic marker that allowed the selection of precise and imprecise excision events was therefore developed in the present study. As previously reported, the LIR is much more unstable in strains with the mutant DNA polymerase δ gene. However, compared with imprecise excision, precise excision accounts for only a small proportion of the genetic changes. Analysis of the imprecise excisants is consistent with a role for replication. The majority of imprecise excisants result from excision between small (4- to 6-bp) repeats in a polar fashion, often leading to the formation of quasipalindrome structures (<300 bp) that subsequently may be highly unstable. Using the newly developed LIR and various derivatives, we demonstrate a novel feature of LIR-induced genetic instability, namely, that LIRs are also hot spots for mitotic interchromosomal recombination.

MATERIALS AND METHODS

Plasmids and strains. To construct the Tn*5*(*URA3*) insert in the yeast *LYS2* gene (diagrammed in Fig. 1), the central *Bgl*II-*Bgl*II 2,784-bp fragment of Tn*5* in the plasmid pACYC184::Tn*5* (17) was replaced by the 1,110-bp *Bgl*II-*Bgl*II *URA3* fragment from the plasmid pFL34 (5). The region removed from Tn*5* contained the unique loop (2,748 bp) and 20 bp of the IS*50* LIR from each side of the loop. The lengths of the remaining LIRs of the Tn*5*(*URA3*) (including *Bgl*II sites) are 1,520 bp. The LIR-*URA3*-LIR fragment was created by cutting Tn*5*(*URA3*) with *Hpa*I, which cuts 185 bp from the external ends of IS*50*. Subsequently, the Tn*5-13* insertion (17) in the chromosomal gene *LYS2* was replaced by Tn*5*(*URA3*) by homologous recombination with the fragment described above, yielding *LYS2*::Tn*5*(*URA3*). The Tn*5*(*URA3*) insert is flanked by 9-bp direct repeats created when the Tn*5* was originally integrated into *LYS2* (18). The Tn*5*(*URA3*) replacements were made in a set of isogenic strains, POL$^+$-DM *MATα lys2*-Tn*5-13 leu2-2 trp1*-Δ*1 ura3-x* (or *ura3*-Δ-*Sma*I-*Hin*dIII deletion of all the *URA3*, leaving only 62 bp of 5′ homology to the *URA3* insert in Tn*5*; the deletion was constructed with the plasmid pJL164, provided by J. Li) and *pol3-t*-DM (same genotypes as POL$^+$-DM, but with *pol3-t*) (16). The replacements were confirmed by genetic and Southern analyses. The point mutation *lys2-111*

* Corresponding author.

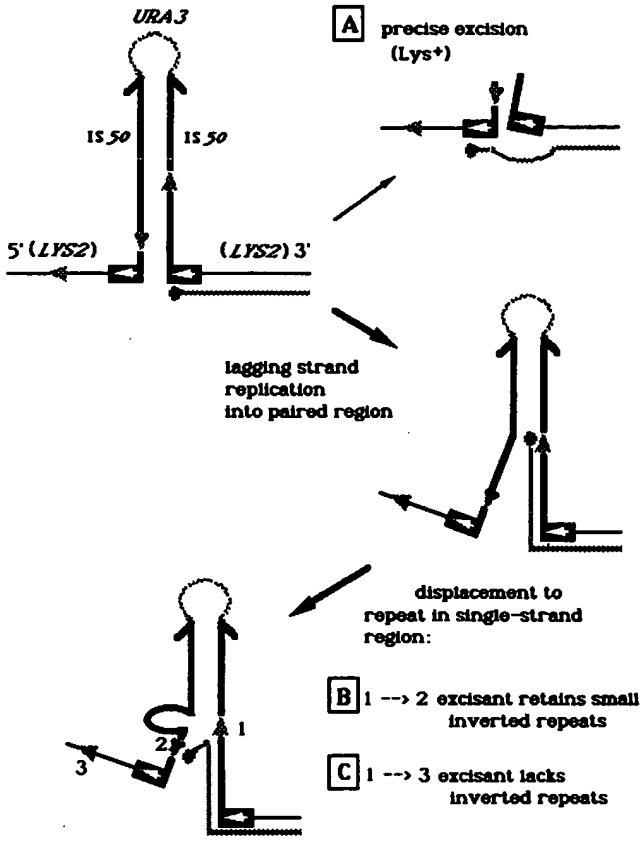


FIG. 1. Insert Tn*5*(*URA3*) in the *LYS2* gene and proposed mechanism for precise and imprecise excision of the insert. The Tn*5*(*URA3*) insert has the central fragment of Tn*5* replaced by the 1,110-bp *URA3* fragment (top left diagram). The lengths of the remaining LIR (thick vertical lines) are 1,520 bp. The Tn*5-13* insertion (17) in the chromosomal gene *LYS2* (thin horizontal lines) was replaced by Tn*5*(*URA3*) via homologous recombination, yielding *LYS2*::Tn*5*(*URA3*). Tn*5*(*URA3*) is flanked by 9-bp direct repeats created when the Tn*5* was originally integrated into *LYS2* (16). Single-strand regions that might appear during replication of the segment containing the LIRs (IS*50* segments [thick vertical arrows]) are proposed to lead to complementary pairing of the LIRs (described in reference 16). We propose that the replicating lagging strand (dotted line with ball corresponding to 3' end) encounters the paired LIR. Slippage between the 9-bp direct repeats (black boxes with arrows) results in precise excision (A). Imprecise excision (B and C) can occur if the lagging-strand replication machinery enters the stem (see the text). Small repeats (dotted arrowheads) are frequent; upon occurrence of a repeat, there is the possibility of slippage to a small repeat in a region of single-strand DNA or of continued replication. If slippage of the newly synthesized strand occurs between small repeats internal to the LIR (i.e., 1→2) (B), the excisant will contain short inverted repeats and may be unstable for Lys$^-$-to-Lys$^+$ reversion, as it can still undergo excision between the two 9-bp direct repeats. If slippage occurs between a small repeat internal to the LIR and an external repeat (i.e., 1→3) (C), there will be no opportunity to generate Lys$^+$ and the excisant will be stable.

(with a G→A base pair change 92 bp in the 5' direction from the Tn*5-13* insertion; the mutation was provided by Y. I. Pavlov) was transferred into the chromosomal *LYS2* allele from the plasmid pLL12-*lys2-111* by gene conversion as described in reference 17.

Genetic and molecular procedures. Genetic and molecular procedures were described previously (16, 17). The *ura3* mutations were isolated on a selective medium containing 5-fluoro-orotic acid (4); a fluctuation test (described in reference 13) was used to determine the mutation rates. Eight or more independent single colonies or groups of up to 10 single colonies grown at either 30 or 20°C (pregrowth) were used for rate determinations. Rates of intragenic recombination of the *lys2* heteroalleles were determined similarly. The *lys2* mutant alleles in the Lys$^+$ recombinants were identified as described in reference 9. For this purpose, the homozygotes for the mutant *lys2* alleles from the independent Lys$^+$ recombinants were selected on a medium containing alpha-aminoadipic acid (7). The ability to recombine with appropriate *lys2* tester mutations was studied to identify the presence of each mutation of the initial pair of heteroalleles. To characterize the genetic stability of the inserts in the *LYS2* genes, frequencies of reversion to Lys$^+$ were determined as described previously (16, 17). To determine the breakpoints of imprecise excision, the region surrounding Tn*5-13* was amplified in 25 cycles (2 min at 94°C, 1 min at 55°C, and 3 min at 72°C) of polymerase chain reaction (PCR) with Vent polymerase (New England Biolabs) according to the manufacturer's instructions with the oligonucleotide primers (ByoSynthesis Inc.) GAGACGCTACCAAGTTCG and CGGCTAAGCTCATAACAT, complementary to the sequences 185 bp 5' and 265 bp 3', respectively, from the ends of Tn*5-13*. The sizes of PCR products were in agreement with results of the Southern analysis. PCR products were sequenced directly (30). Formamide (40%) was included in the sequencing gels when needed to resolve compression artifacts.

RESULTS

Deletions associated with LIRs. Previously, Gordenin et al. demonstrated that Tn*5* as a model LIR was excised at a high rate in DNA polymerase-deficient mutants (16). However, only reversions to Lys$^+$ through precise excision (breakpoints in the duplicated 9 bp of the Tn*5* target) or unique in-frame imprecise excisions associated with LIRs were detectable in this study. In order to understand the general consequences of LIRs, we developed a genetic system in which the complete spectrum of Tn*5* deletions could be selected. For this purpose, a 3-kb internal region of Tn*5* was replaced by *URA3* [i.e., *LYS2*::Tn*5*(*URA3*)] (diagrammed in Fig. 1). By selecting *ura3* auxotrophs on 5-fluoro-orotic acid (4), a broad spectrum of genetic changes, including both precise (Ura$^-$ Lys$^+$) and imprecise (Ura$^-$ Lys$^-$) excisants, could be identified. The Ura$^-$ Lys$^-$ category included mutations specific to the *URA3* gene. The Ura$^-$ Lys$^-$ auxotrophs could be categorized according to their ability to revert to Lys$^+$, a feature allowing discrimination between excisants (described below).

A defect in replicative polymerase δ resulted in a 45-fold-higher rate of appearance of *ura3* mutants at the semipermissive temperature (30°C) than at the permissive temperature (Table 1). This contrasted with the low rate in the Pol$^+$ strain at 30°C. Only a small percentage of Ura$^-$ isolates were due to precise excision, i.e., were Ura$^-$ Lys$^+$. As discussed below, nearly half of the events at the semipermissive temperature of growth for the *pol3* mutant were due to imprecise excision.

The Ura$^-$ Lys$^-$ mutants were genetically characterized according to their ability to revert to Lys$^+$ (the result of the precise excision of the remaining insert). For the *pol3-t* strain, there were three categories of Ura$^-$ Lys$^-$ mutants (Table 1). The first (with *ura3* mutations) yielded high levels

TABLE 1. Frequencies and categories of Ura⁻ mutants arising from the insert *LYS2*::Tn*5*(*URA3*) in Pol⁺ and *pol3-t* strains[a]

Strain and pregrowth temp	Ura⁻ mutation rate (10^6)	% of Ura⁻ isolates (no. of events) generated by:			
		ura3 mutation	Precise excision (Lys⁺)	Imprecise excision[b]	
				Stable	Unstable
pol3-t					
20°C	1.4	78 (72)	1.0 (1)	7.0 (6)	14 (13)
30°C	63	52 (109)	1.4 (3)	4.7 (10)	42 (90)
Pol⁺					
30°C	0.12	95 (55)	1.7 (1)	0 (0)	3.3 (2)

[a] The *pol3-t* and Pol⁺ strains are the derivatives of *pol3-t*-DM and POL⁺-DM strains with the *LYS2*::Tn*5*(*URA3*) insertion. The *ura3* mutations were obtained on a selective medium containing 5-fluoro-orotic acid (4); a fluctuation test (described in reference 13) was used to determine the mutation rates. Since the rates for the strains carrying *ura3-x* and those carrying the *ura3* deletion did not differ significantly, the results were combined. The characterization of the reversion to Lys⁺ (stability) was made with the mutants obtained in the *ura3-x* mutant strains.

[b] Stable refers to the ability to revert from Lys⁻ to Lys⁺. Stable Ura⁻ Lys⁻ isolates exhibited little ($<10^{-6}$) or no reversion to Lys⁺ at 20 or 30°C, while unstable Ura⁻ Lys⁻ isolates exhibited a high frequency (10^{-5} to 10^{-6}) of Lys⁻ to Lys⁺ reversion at both 20 and 30°C.

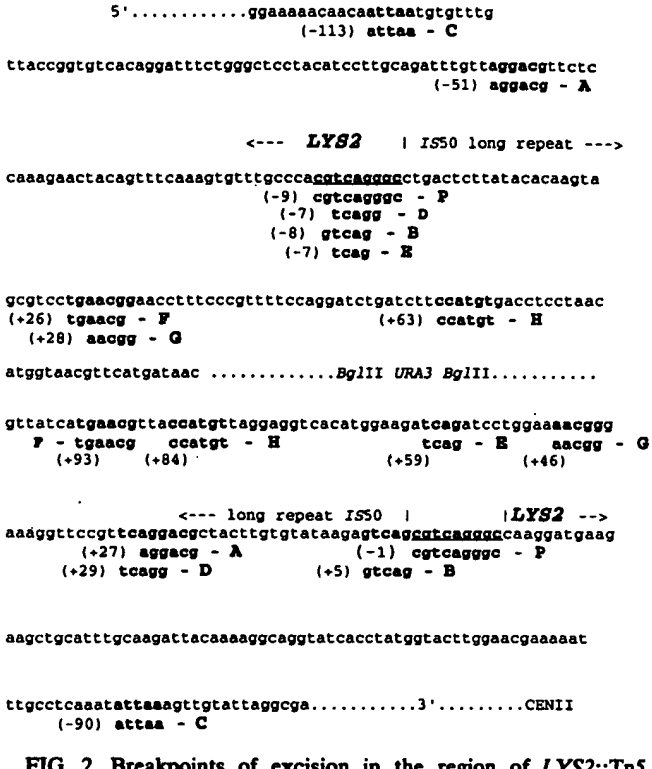


FIG. 2. Breakpoints of excision in the region of *LYS2*::Tn*5* (*URA3*). The Tn*5*(*URA3*) insertion into *LYS2* (including the IS*50* long repeats) (Fig. 1) is located between the 9-bp repeats (18) (underlined sequence). Included in this diagram are 100 bp of each IS*50* (2) and 100 bp of *LYS2* (23) to the left and right of the insert. The orientation of *LYS2* in chromosome II is based on unpublished results of Gordenin et al. (15a). Eight pairs of breakpoints (boldface) corresponding to imprecise excisions are designated A through H. Breakpoints C to H were sequenced in this work. Breakpoints A and B were identified previously and corresponded to in-frame imprecise excisions of Tn*5-13* selected as Lys⁺ (16). Precise excision breakpoints are identified by P. The nucleotide coordinates in the left and right parts of *LYS2* start from the border of the repeats with IS*50* and are given negative values beginning at the start of the repeat. IS*50* nucleotides are given positive values.

of Lys⁺ only at 30°C, which was similar to previous results with Tn*5* (16, 17). These Ura⁻ mutants were shown by Southern analysis to be due to mutations in *URA3* (among 16 mutants in the *ura3-x* strain, all appeared to be point or, at most, small deletions, as determined by gel electrophoresis analysis) (data not shown). The second genetic category of Ura⁻ Lys⁻ isolates exhibited a phenotype which was novel in that there was a high frequency (10^{-5} to 10^{-6}) of Lys⁻ to Lys⁺ reversion at both 20 and 30°C; because of this property, this category was referred to as unstable. A third genetic category of Ura⁻ Lys⁻ mutants consisted of isolates, designated as stable, that exhibited little ($<10^{-6}$) or no reversion to Lys⁺ at either temperature. On the basis of genetic and molecular analysis (described below), we determined that these last two categories are due to imprecise excision.

The spectrum of changes for the Pol⁺ strain was different from that for the *pol3-t* strain in that nearly all the Ura⁻ mutants (23 of 24) that exhibited the observed low rate of Lys⁺ reversion carried mutations in the *URA3* gene, as determined by Southern analysis of the mutants obtained in *ura3-Δ* strain (data not shown). This is consistent with the observation that Tn*5* excision is low in Pol⁺ strains (16). The lower rates of *URA3* mutation in Pol⁺ versus *pol3-t* strains are consistent with the previously reported mutator phenotype of the temperature-sensitive polymerase δ mutants (17, 19, 34). (In an analysis of spontaneous mutations arising in the *pol3* mutant *mut7-1*, deletions [<70 bp] between small repeats accounted for nearly 20% of the mutants [34].) Ectopic gene conversion of the *ura3-x* point mutation cannot be excluded as a source of at least some of the Ura⁻ mutants. The rates of appearance of Ura⁻ mutants did not differ (see the footnotes for Table 1) for strains having a point mutation or a deletion at the normal location of *URA3*; therefore, it is unlikely that ectopic conversion is the major source of point mutations in *URA3* between the LIRs.

The high rate of appearance of the Ura⁻ Lys⁻ mutants in the *pol3-t* strains that were unstable (at 20 and 30°C) or stable for reversion to Lys⁺ was due to imprecise excision. As determined by Southern blot and/or PCR analysis of the three stable and nine unstable excisants, they retained small remnants from the original Tn*5*(*URA3*) that were less than 300 bp. An additional stable excisant was due to a deletion that encompassed all of Tn*5*(*URA3*) plus 203 bp of the surrounding *LYS2* sequence (excisant C; see Discussion). Therefore, all breakpoints are clustered near the ends of the 4,150-bp Tn*5*(*URA3*) insert. Sequence analysis was carried out on 9 of the 13 excisants (4 stable and 5 unstable), and six categories of deletions based on breakpoints were identified (Fig. 2). All breakpoints have short (4- to 6-bp) repeats. There were three excisants in the *ins-G* category and two excisants in the *ins-H* category. These results agree with previous analyses (16) of in-frame imprecise (included in Fig. 2) and precise Tn*5* excision in *S. cerevisiae*. All excisants independently isolated in this study, except excisant C, exhibited a polarity of breakpoints (see Discussion).

Instability of quasipalindrome remnants. The unstable (Lys⁻ to Lys⁺) excisants were caused by breakpoints within the IS*50* repeats of the LIR (Fig. 1), leading to small quasipalindrome structures. Reversion to Lys⁺ occurred primarily through precise excision between 9-bp direct re-

peats at the base of Tn*5*. Previously, Gordenin et al. demonstrated that a selective medium containing alpha-aminoadipic acid (7) can be used to distinguish precise excision from in-frame imprecise excision of *LYS2*::Tn*5-13* (16). Precise excision generates the wild-type gene, resulting in sensitivity to alpha-aminoadipic acid (Adp^-), whereas cells containing the pseudo-wild-type gene may be capable of growth on this inhibitor (Adp^+) (16). On the basis of the low fraction of Adp^+ revertants, we conclude that reversion to Lys^+ of the unstable as well as stable excisants described in Table 2 was accomplished primarily by interactions between the direct repeats originally at the base of Tn*5* (i.e., precise excision). Among revertants of the stable excisant E, a large fraction of the Adp^+ isolates appear to reflect another mechanism of reversion (possibly imprecise excision or suppression).

The excision frequency of quasipalindromes appears to be related to the length of DNA between the inverted repeats, as demonstrated by excisants F and G. The lengths of the inverted repeat regions are comparable; however, the intervening loop sequences are 8 bp for the highly unstable structure, compared with 55 bp for the more stable structure. The unstable excisant H also contains a small loop sequence (9 bp) between inverted repeats. Since the number of quasipalindrome inserts examined is small, additional comparable inserts need to be investigated to confirm the effect of the intervening loop on stability. Three other unstable remnants (reversion frequencies of about 10^{-5}) that are approximately 300 bp long with inverted repeats greater than 90 bp per repeat have been identified (data not shown).

The unstable quasipalindrome structures isolated in the *pol3-t* strain are also unstable in the Pol^+ strain, although the reversion frequencies are an order of magnitude less. Unlike Tn*5*(*URA3*) and Tn*5*, they revert at a high frequency at the permissive temperature in the *pol3-t* mutant.

Three rare reverting excisants (C, D, and E [Fig. 2 and Table 2]) do not contain any inverted repeats and are due to a deletion of one or both IS*50* sequences of the LIR. The reversion frequencies are much lower in Pol^+ than in the *pol3-t* strain. Possibly the revertants correspond to rare replication slippage, as indicated by the isolation of deletions between closely spaced small direct repeats in another *pol3* mutant (34). If so, the present results indicate that the *pol3-t* mutation can have an effect on replication by polymerase δ, even at the permissive temperature (also see Table 1).

LIRs create a hotspot for mitotic recombination. As we have shown, LIRs exhibit unique properties in that they can lead to intrachromosomal instabilities. The proposed replication mechanism(s) (see reference 16 and Discussion) suggests that LIRs could also influence interchromosomal recombination. We therefore measured heteroallelic recombination (gene conversion) between the point mutation *lys2-8* (9, 26) located approximately 0.6 to 0.8 kb upstream (5′) from Tn*5* and various *lys2* insertions. Among the latter were the LIR insertions Tn*5-13* and Tn*5*(*URA3*), as well as the 24-bp (*ins-D*) and 55-bp (*ins-E*) pieces of the external end of IS*50* and quasipalindromes *ins-G* and *ins-H* (Table 2 and Fig. 2); all insertions were at the same position. We also studied mitotic recombination of *lys2-8* with the point mutation *lys2-111* located 92 bp from Tn*5-13* in the direction of the *lys2-8* (see Materials and Methods). Reversion rates of the heteroalleles are shown in Table 3. On the basis of the high reversion rates, we conclude that the major source of reversion is heteroallelic recombination. The frequency of Lys^+ for all heteroallelic diploids was 10 to 10^5 times more than the reversion frequency of the individual *lys2* mutations (Table 2; data for *lys2-8* are not shown). The reversion rates of Tn*5-13* in the homozygous *pol3* and Pol^+ diploids did not differ significantly from the rates in the haploids (data not shown).

In the *pol3* mutant, recombination was generally higher at the semipermissive than at the permissive temperature for all pairs of heteroalleles tested, as expected from previous reports of the effects of polymerase mutations on recombination (1) (Table 3). However, when one of the heteroalleles was the LIR insert, the recombination rate was 4- to 18-fold higher than that for the other heteroalleles in both the *pol3* and Pol^+ isogenic strains (at 20 as well as 30°C). The increase in recombination frequencies is accompanied by a high (up to sevenfold) bias in favor of loss of the Tn*5* allele compared with the loss of *lys2-8* (Table 4). The bias is primarily due to a preference in gene conversion of the Tn*5* allele. (See reference 8 for a discussion of genetic outcomes of gene conversion and reciprocal recombination in mitotic cells.) We conclude that the observed induction and pattern of mitotic homologous recombination are due specifically to the LIRs (see Discussion).

DISCUSSION

Inverted repeats are a frequent source of genetic instability in bacteria. Using Tn*5* as a source of LIRs and using newly developed derivatives that allow the detection of a broad range of changes, we have examined the consequences of various types of inverted repeats on genomic stability in yeast. Three types of genetic instabilities associated with inverted repeats have been demonstrated: a high rate of large deletions in a DNA polymerase δ mutant with the breakpoints clustering near the ends of LIRs, the stimulation of interchromosomal mitotic recombination in Pol^+ and *pol3* mutants, and the frequent deletion of quasipalindromes.

The increased frequency of deletions in a *pol3* mutant and the clustering of breakpoints support the idea that altered replication of LIRs could be the source of deletions. An analysis of the breakpoints of the imprecise excision led us to conclude that the inverted repeats can cause excision via short repeats either internal or external to the LIRs. The regions that include all the sequenced breakpoints (i.e., 100 bp on either side of Tn*5* ends [Fig. 2]) contain 128 short (4- to 8-bp) homologous sequences between them. The much higher frequency of imprecise excision than of precise excision (20- to 30-fold in the *pol3-t* strain) may be a reflection of the large number of small repeats available for imprecise excision.

There is polarity of excision such that a breakpoint in one IS*50* repeat of the LIR [portrayed as the right repeat of *LYS2*::Tn*5*(*URA3*) in Fig. 1] is always higher than the breakpoint in the other IS*50* repeat (except for excisant C, in which the breakpoints are external, yet near the LIRs). Previous results with the imprecise excisants A and B (16) are consistent with the observed polarity, although there may be selection associated with these in-frame deletions. Further support for the polarity rule comes from a consideration of the orientation of short repeats within an LIR. That is, for any short sequence within one of the long repeats of the LIR, there is a corresponding short sequence of the opposite orientation in the other long repeat. For any pair of short direct repeats in an LIR, there is a corresponding symmetrical pair of short direct repeats. While there are two symmetrical opportunities for polarity of breakpoints, only one was observed. We propose that the polarity is a reflec-

TABLE 2. Reversion rates of insertions at the *LYS2*::Tn*5-13* site

Insertion (type of structure)	No. of isolates[a]	Remaining structure (bp)	Rate of reversion to Lys$^+$ (10^6)[b]			
			pol3-t		Pol$^+$	
			20°C	30°C	20°C	30°C
ins-D (stable D)	1	24 / 7 7	0.2 ± 0.02 (4/120)[c]	0.9 ± 0.07 (1/128)	0.008 ± 0.001 (4/25)	0.018 ± 0.007 (0/33)
ins-E (stable E)	1	55 / 6 6	0.1 ± 0.07 (18/120)	0.3 ± 0.03 (15/115)	0.007 ± 0.001 (4/27)	0.004 ± 0.001 (5/8)
ins-F (stable F)	1	32 55 32 / 9 9	0.4 ± 0.02 (9/115)	0.4 ± 0.03 (1/128)	0.03 ± 0.004 (0/42)	0.004 ± 0.001 (0/22)
ins-G (unstable G)	3	33 8 33 / 9 9	3.5 ± 0.1 (2/348)	6.8 ± 0.3 (8/358)	0.5 ± 0.05 (1/112)	0.6 ± 0.06 (0/107)
ins-H (unstable H)	2	69 9 69 / 9 9	3.6 ± 0.2 (10/235)	4.9 ± 0.2 (18/197)	0.3 ± 0.016 (3/111)	0.25 ± 0.02 (3/116)
Tn*5*(*URA3*) (DM)	NA	1520 1100 1520 / 9 9	0.016 ± .002 (1/70)	1.8 ± 0.12 (8/160)	0.0003 (NT)	0.002 (NT)

[a] The independent isolates were K31 (C), K25 (D), K27 (E), K26 (F), K14 (G), K15 (G), K23 (G), K10 (H), and K13 (H). For the Pol$^+$ strain all isolates except K13, K15, and K23 were used. NA, not applicable.

[b] No Lys$^+$ revertants were obtained for excisant C ($<10^{-9}$). Therefore, this was not included in the table. The *pol3-t*-DM strains containing excisants D through H (also described for Fig. 2) were transformed with the *POL3* gene (16). Four transformants of each type of excisant strain were used to determine the frequencies of Lys$^-$ to Lys$^+$ reversion. Presented are the rates for all Lys$^+$ revertants which include those that are also Adp$^+$ (grow on selective medium containing alpha-aminoadipate). The average values for 12 or more experiments ± standard errors are given for all except two cases in which the frequencies were low. The rates determined in separate fluctuation tests (not shown) with the *pol3-t* excisants G and H were close to the values shown.

[c] Numbers in parentheses are numbers of Lys$^+$ Adp$^+$ revertants/total number of Lys$^+$ revertants tested. NT, not tested.

TABLE 3. Influence of the Tn*5* LIRs on mitotic homologous recombination

lys2 heteroalleles[a]	Recombination rate (10^6)			
	pol3-t		Pol$^+$	
	20°C	30°C	20°C	30°C
Tn*5-13*/-8	30	406	45	64
Tn*5*(*URA3*)/-8	NT[b]	NT	56	44
ins-D/-8	4.5	36	3.1	8.9
ins-E/-8	5.1	50	5.5	8.4
ins-G/-8	NT	NT	NT	4.6
ins-H/-8	NT	NT	NT	9.0
111/-8	4.0	32	NT	16

[a] The isogenic *pol3-t* and Pol$^+$ strains containing the inserts Tn*5-13*, Tn*5*(*URA3*), *ins-D*, *ins-E*, *ins-G*, *ins-H* (see the legend to Fig. 1 and the footnotes to Table 2), or the point mutation *lys2-111* (see Materials and Methods) were crossed with the POL$^+$-DA strain (*MATα lys2-8 his3-15,11 ura3 leu2 trp1-del1*) or the *pol3-t*-DA strain (the congenic *pol3-t* strain with the same auxotrophic markers). The Pol$^+$ strain was homozygous for the *POL3* gene. All resulting diploids were heteroallelic for the *LYS2* mutations and, therefore, auxotrophic for lysine. The Lys$^+$ intragenic recombinants were selected on medium with no lysine. The rates of heteroallelic recombination were determined by fluctuation testing of 10 to 15 cultures. Independent frequency measurements (16, 17) confirmed the differences (data not shown).
[b] NT, not tested.

tion of the direction of replication of the lagging strand. The polarity rule for breakpoints is predicted by the model of Tn*5* excision via slipped mispairing in the replicating lagging strand (3, 16). (While polymerase δ is proposed as a lagging-strand polymerase [24], we note that our model could apply even if the polymerase functioned both in leading- and lagging-strand replication.) According to the model, single-stranded regions arising during altered lagging-strand replication of Tn*5* can form a secondary stem-and-loop structure within the LIR, thereby delaying or blocking subsequent DNA synthesis. We propose that although lagging-strand DNA synthesis is generally not considered to involve strand displacement, the lagging-strand replication machinery may enter the duplex formed by the internal repeats (Fig. 1). However, as it replicates into the stem, the replicating end may be unstable because of this uncommon form of lagging-strand replication. As a result, opportunities to undergo replication on a single strand may be favored. Replication into the stem would result in one IS*50* repeat of the LIR becoming duplex and the other becoming single stranded. A deletion event could occur by strand slippage from the duplex IS*50* repeat of the LIR to a repeat in the single-strand region. Since slippage into a single-strand region (rather than into an already-paired region) is likely to occur, polarity in relation to an origin of replication is predicted. For the region studied, the present results suggest that the *LYS2* gene is replicated from the chromosomal origins(s) located on the 5′ side of *LYS2* (centromere distal).

Polarity, along with the observations on imprecise excision above, also argues against LIR deletions being strictly the consequence of altered replication (34) rather than the consequence of interaction of the replication machinery with the LIR. The influence of replication direction on palindrome-stimulated deletions within a plasmid was recently demonstrated for *Escherichia coli* (31).

The observations that altered replication could enhance recombination (1) and that altered replication of LIRs can yield deletions led us to investigate whether LIRs could also induce interchromosomal recombination. We found that LIRs create a hotspot for mitotic intragenic recombination in *pol3-t* and Pol$^+$ isogenic strains, in which mitotic gene conversion preferentially results in the loss of the LIR-containing insert (Tn*5*). While *pol3-t* mutants exhibit enhanced recombination for all heteroallelic pairs, there is still a bias in favor of the LIR-containing insert. It remains to be established whether the bias will be seen with other mutants or agents that induce recombination. Several observations have led us to conclude that the enhanced recombination rates and conversion bias are specifically due to the LIRs. All comparisons were made between isogenic strains. All pairs of mutant alleles except one were at the same positions in the *LYS2* gene. The exception (point mutation *lys2-111*) was also located close to the site of Tn*5* insertion. The enhanced recombination is not due to the sequences near the ends of the LIRs, judging by the lack of increase by small inserts corresponding to one or both ends of the LIRs. The enhanced recombination rates and the conversion bias cannot be attributed simply to a nonspecific effect of a large insertion. We and others (9, 32) have demonstrated with several heteroallelic pairs of mutations in the *LYS2*, *URA3*, *HIS4*, and *HIS3* genes that a large insertion (as well as deletions) can lead to as much as a 19-fold bias in favor of loss of a point mutation; this is in contrast to the present findings involving the Tn*5* insertion. (The only exception was *lys2-32* [9]; however, this mutation was subsequently shown to be due to a large Ty1 insertion [8].) The absence of conversion bias for the heteroallelic pairs of point mutations that do not exhibit hotspots is well documented (28). A conversion bias involving a hotspot has been reported for mitotic gene conversion of the *HOT1* allele of yeast (33).

In terms of the model described above, these results suggest that homologous recombination may be an alternative to replication slippage. Possibly, the growing 3′ end, when stalled at the base of a LIR secondary structure (Fig. 1), can interact with another homologous sequence, resulting in gene conversion. This is reminiscent of the mechanism

TABLE 4. Genotypes of Lys$^+$ recombinants resulting from recombination between the heteroalleles with or without LIRs[a]

lys2 heteroalleles	No. of recombinants								
	pol3-t						Pol$^+$ (30°C)		
	20°C			30°C					
	+/8	+/N	+/8N	+/8	+/N	+/8N	+/8	+/N	+/8N
Tn*5-13*/-8	106	23	1	103	15	1	103	31	7
ins-E/-8	NT	NT	NT	55	44	5	100	98	7
111/-8	67	78	1	40	39	16	52	61	3

[a] Genotypes of the Lys$^+$ recombinants were determined as described in Materials and Methods and reference 9. N, insertion of Tn*5-13*, *ins-E* alleles, or point mutation *lys2-111*; 8, *lys2-8* allele; 8N indicates the presence of both alleles in the mutant *lys2* gene. nt, not tested.

proposed for DNA synthesis past a lesion in phage T4 (15). If this model is correct, the LIRs themselves could lead to ectopic recombination with another copy of the same repeat elsewhere in the genome when the 3′ end is stalled after entering the paired region (Fig. 1). This possibility is currently under investigation. As an alternative explanation for the observed recombinogenesis of LIRs, it is also possible that the repeats within a LIR have an increased probability of recombinational interaction; this in turn could result in a greater likelihood of interchromosomal recombination. The absence of stimulated recombination by quasipalindromes *ins-G* and *ins-H* could reflect a greater likelihood of bypassing the secondary structure formed by the short quasipalindrome in the course of replication than of bypassing the LIR secondary structure. This is consistent with the higher deletion frequencies of the quasipalindromes.

Palindromes and the unstable quasipalindromes described here are a special class of inverted repeats. High frequencies of palindrome excision have been observed in bacteria (references 10, 11, and 14 and references therein) and recently in *S. cerevisiae* (20, 29). Unlike Tn*5* and Tn*5* (*URA3*), the unstable quasipalindrome structures isolated in the *pol3-t* strain are unstable in the Pol$^+$ strain, and they revert at a high frequency at the permissive temperature in the *pol3-t* mutant. Possibly because of their small size, they can lead to secondary structures during normal replication, whereas intrastrand duplexes may occur between the LIRs of Tn*5* only in mutants with altered replication (Fig. 1) (16). Small palindromes in bacterial plasmids have recently been demonstrated to generate stem-and-loop fragments physically separate from the parental plasmid, suggesting template switching during replication as a source of excision (27). If unstable palindromes in eukaryotes have the same properties, such a mechanism could be a source of random, albeit rare, insertions into the DNAs of higher organisms.

These results describing the multiple effects of inverted repeats on genome stability suggest that LIRs may be an important source of genetic change in genomes that contain many repeats. For example, within the human genome there are on the order of a million repeats that are greater than a few hundred base pairs (12, 21). Assuming randomness of orientation, the number of LIRs is considerable and the distances between them are comparable to those described in the present report. If the opportunities for interactions extend over large distances, the number of possible interactions becomes immense. Such interactions may explain the instability of human DNAs in yeasts (22a, 25). Given that there are comparable DNA polymerases in eukaryotes (6, 35), we propose that the human genome is at considerable risk for rearrangements due to deletions between LIRs or through LIR-stimulated recombination. The extent of LIR-induced changes is expected to relate to the level of homology between repeats. The consequences of DNA divergence are currently being investigated.

ACKNOWLEDGMENTS

We are grateful to L. Golovanova, N. Koloteva, and V. Hashem for assistance in experiments, to K. Tindall and Z. Kelesceny for advice and aid in sequencing procedures with PCR, to Y. I. Pavlov for providing the *lys2-111* mutant, to J. Li for providing the plasmid pJL164, and to R. C. von Borstel and A. Morrison for critical comments.

Support was provided for different parts of this work by the Russian National Program, Frontiers in Genetics (A.M.), by the Russian Foundation for Basic Research grant 93-04-6032 (D.A.G.), and by the Fogarty Center Central and Eastern European Initiative (M.A.R. and D.A.G.). E.P. was supported by a National Institutes of Health postdoctoral fellowship.

REFERENCES

1. **Aguilera, A., and H. Klein.** 1988. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*: I. Isolation and genetic characterization of hyper-recombination mutations. Genetics **119**:779–790.
2. **Auerswald, E.-A., G. Ludwig, and H. Schaller.** 1981. Structural analysis of Tn*5*. Cold Spring Harbor Symp. Quant. Biol. **45**:107–113.
3. **Berg, D. E.** 1989. Transposon Tn*5*, p. 185–210. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
4. **Boeke, J. D., F. LaCroute, and G. R. Fink.** 1984. A positive selection for mutants lacking orotidine-5-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. **197**:345–346.
5. **Bonneaud, N., O. Ozier-Kalogeropulos, L. Gy, M. Labouesse, L. Minvielle-Sebastia, and F. LaCroute.** 1991. A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae/E. coli* shuttle vectors. Yeast **7**:609–615.
6. **Burgers, P. M. J., R. A. Bambara, J. L. Campbell, L. M. S. Chang, K. M. Downey, U. Hubscher, M. Y. W. T. Lee, S. M. Linn, A. G. So, and S. Spadar.** 1990. Revised nomenclature for eukaryotic DNA polymerases. Eur. J. Biochem. **191**:616–618.
7. **Chattoo, B. B., F. Sherman, D. A. Azubalis, T. Fjelstedt, D. Mehnert, and M. Ogur.** 1979. Selection of *lys2* mutants of yeast *S. cerevisiae* by utilization of alpha-aminoadipate. Genetics **93**:51–65.
8. **Chernoff, Y. O., and D. A. Gordenin.** 1987. The phenotypic expression of yeast transposon insertion at *LYS2* gene and of deletions resulting from imprecise excision. Genetika **23**:30–40. (English translation.)
9. **Chernoff, Y. O., O. V. Kidgotko, O. Demberelijn, I. L. Luchnikova, S. P. Soldatov, V. M. Glaser, and D. A. Gordenin.** 1984. Mitotic intragenic recombination in yeast *Saccharomyces*: marker effects on conversion and reciprocity of recombination. Curr. Genet. **9**:31–37.
10. **Collins, J.** 1980. The instability of palindromic DNA. Cold Spring Harbor Symp. Quant. Biol. **45**:409–416.
11. **Collins, J., G. Volckaert, and P. Nevers.** 1982. Precise and nearly precise excision of the symmetrical inverted repeats of Tn*5*: common features of *rec*A-independent deletion events in *E. coli*. Gene **19**:139–146.
12. **Deininger, P. L.** 1989. SINEs: short interspersed repeated DNA elements in higher eucaryotes, p. 619–636. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
13. **Drake, J. W.** 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88**:7160–7164.
14. **Ehrlich, D. S.** 1989. Illegitimate recombination in bacteria, p. 799–832. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
15. **Formosa, T., and B. Alberts.** 1986. DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins. Cell **47**:793–806.

15a. **Gordenin, D. A., et al.** Unpublished data.

16. **Gordenin, D. A., A. L. Malkova, A. Peterzen, V. N. Kulikov, Y. I. Pavlov, E. Perkins, and M. A. Resnick.** 1992. Transposon Tn*5* excision in yeast: the influence of DNA polymerases alpha, delta, epsilon and repair genes. Proc. Natl. Acad. Sci. USA **89**:3785–3789.
17. **Gordenin, D. A., Y. Y. Proscyavichus, A. L. Malkova, M. V. Trofimova, and A. Peterzen.** 1991. Yeast mutants with increased transposon Tn*5* excision. Yeast **7**:37–50.
18. **Gordenin, D. A., M. V. Trofimova, O. N. Shaburova, Y. I. Pavlov, Y. O. Chernoff, Y. V. Chekuolene, Y. Y. Proscyavichus, K. V. Sasnauskas, and A. A. Janulaitis.** 1988. Precise excision of bacterial transposon Tn*5* in yeast. Mol. Gen. Genet. **213**:388–393.
19. **Hartwell, L., and D. Smith.** 1985. Altered fidelity of mitotic

chromosome transmission in cell cycle mutants of *S. cerevisiae*. Genetics **110**:381–395.
20. **Henderson, S. T., and T. D. Petes.** 1993. Instability of a plasmid borne inverted repeat in *Saccharomyces cerevisiae*. Genetics **133**:57–62.
21. **Hutchison, C. A., III, S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell.** 1989. LINEs and related retroposons: long interspersed sequences in the eucaryotic genome, p. 593–617. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
22. **Kleckner, N.** 1989. Transposon Tn*10*, p. 227–268. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
22a. **Larionov, V. L., et al.** Submitted for publication.
23. **Moris, M., and S. Jinks-Robertson.** 1991. Nucleotide sequence of the *LYS2* gene of *S. cerevisiae*: homology to *B. brevis* tyrocidine synthetase 1. Gene **98**:141–145.
24. **Morrison, A., H. Araki, A. B. Clark, R. K. Hamatake, and A. Sugino.** 1990. A third essential DNA polymerase in *S. cerevisiae*. Cell **62**:1143–1151.
25. **Neil, D. L., A. Villasante, R. B. Fisher, D. Vetrie, B. Cox, and C. Tyler-Smith.** 1990. Structural instability of human tandemly repeated DNA sequences cloned in yeast artificial chromosome vectors. Nucleic Acids Res. **18**:1421–1428.
26. **Noskov, V. N., M. G. Tarutina, Y. I. Pavlov, V. N. Kulikov, M. V. Trofimova, A. V. Gorbacheva, Y. O. Chernoff, K. V. Sasnauskas, M. A. Neistat, I. I. Tolstorukov, and D. A. Gordenin.** 1990. Development of intragenic mapping system for molecular genetic analysis of mutations in *LYS2* gene of *S. cerevisiae*. Genetika **26**:1161–1168. (English translation.)
27. **Ohshima, A., S. Inouye, and M. Inouye.** 1992. *In vivo* duplication of genetic elements by the formation of stem-loop DNA without an RNA intermediate. Proc. Natl. Acad. Sci. USA **89**:1016–1020.
28. **Petes, T. D., R. E. Malone, and L. S. Symington.** 1991. Recombination in yeast, p. 407–521. *In* J. R. Broach, E. W. Jones, and J. R. Pringle (ed.), The molecular and cellular biology of the yeast saccharomyces: genome dynamics, protein synthesis, and energetics. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
29. **Ruskin, B., and G. R. Fink.** 1993. Mutations in *POL1* increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*. Genetics **133**:43–56.
30. **Tindall, K. R., and L. F. Stankowski.** 1989. Molecular analysis of spontaneous mutations at the gpt locus in Chinese hamster cells. Mutat. Res. **220**:241–253.
31. **Trinh, T. Q., and R. R. Sinden.** 1991. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. Nature (London) **352**:544–547.
32. **Vincent, A., and T. D. Petes.** 1989. Mitotic and meiotic gene conversion of Ty elements and other insertions in *Saccharomyces cerevisiae*. Genetics **122**:759–772.
33. **Voelkel-Meiman, K., and G. S. Roeder.** 1991. A chromosome containing *HOT1* preferentially receives information during mitotic interchromosomal gene conversion. Genetics **124**:561–572.
34. **von Borstel, R. C., R. Ord, S. P. Steward, R. G. Ritzel, G. S.-F. Lee, U. G. G. Hennig, and E. A. Savage.** The mutator *mut7-1* of *Saccharomyces cerevisiae*. Mutat. Res., in press.
35. **Wang, T. S.-F.** 1991. Eukaryotic DNA polymerases. Annu. Rev. Biochem. **60**:513–552.

JOURNAL OF VIROLOGY, July 1996, p. 4805–4810
0022-538X/96/$04.00+0

Vol. 70, No. 7

Efficient Generation of Recombinant Adenovirus Vectors by Homologous Recombination in *Escherichia coli*

C. CHARTIER, E. DEGRYSE, M. GANTZER, A. DIETERLÉ, A. PAVIRANI, AND M. MEHTALI*

Gene Therapy Department, Transgène S.A., 67000 Strasbourg, France

Received 18 January 1996/Accepted 12 April 1996

Despite recent technical improvements, the construction of recombinant adenovirus vectors remains a time-consuming procedure which requires extensive manipulations of the viral genome in both *Escherichia coli* and eukaryotic cells. This report describes a novel system based on the cloning and manipulation of the full-length adenovirus genome as a stable plasmid in *E. coli*, by using the bacterial homologous recombination machinery. The efficiency and flexibility of the method are illustrated by the cloning of the wild-type adenovirus type 5 genome, the insertion of a constitutive promoter upstream from the E3 region, the replacement of the E1 region by an exogenous expression cassette, and the deletion of the E1 region. All recombinant viral DNAs were shown to be fully infectious in permissive cells, and the modified E3 region or the inserted foreign gene was correctly expressed in the infected cells.

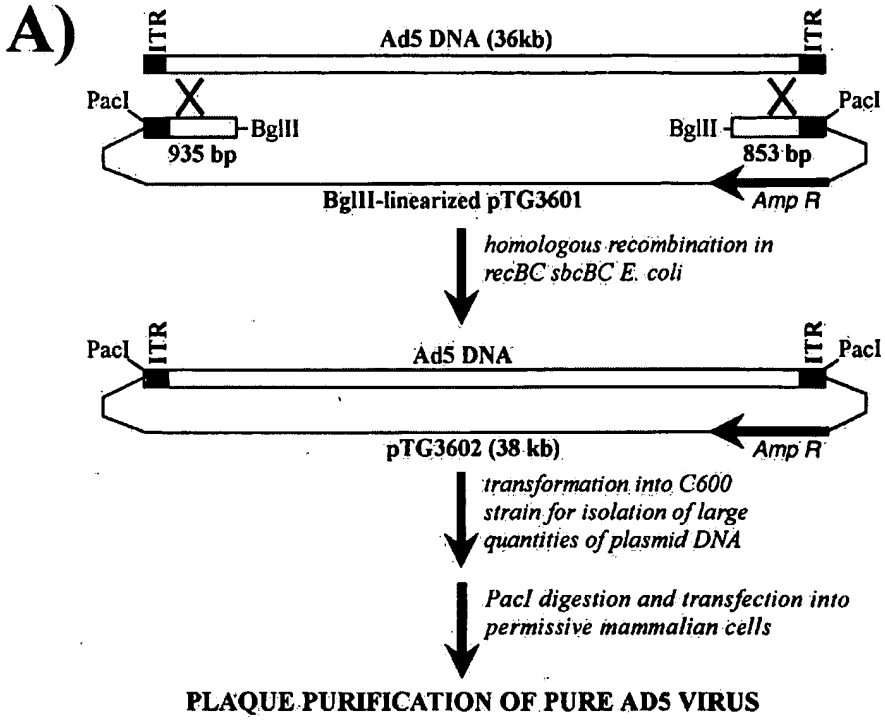
Adenoviruses are generally associated with benign pathologies in humans and are characterized by a number of features which make them particularly attractive as gene transfer vectors for gene therapy or immunization purposes (7, 15, 21, 35). Most vectors are based on the adenovirus type 5 (Ad5) backbone in which an expression cassette containing the foreign gene has been introduced in place of the early region 1 (E1) or early region 3 (E3). Viruses in which E1 has been deleted are defective for replication and are propagated in the human 293 complementation cells, providing the E1 products in *trans* (16). Methods to construct such recombinant adenoviruses are well documented (3, 14). They usually consist of a first step of subcloning of the exogenous expression cassette into a segment of the viral genome. The recombinant vector is then produced by the reconstitution of a complete viral DNA molecule through ligation in vitro between the segment and the viral genome, followed by transfection into permissive cells. Alternatively, cotransfection into the complementation cells of the viral genome and plasmid DNA molecules can generate the recombinant viruses by homologous recombination in vivo. These methods frequently generate a background of nonrecombinant viruses, and despite recent improvements (23), repeated screening of many plaques is sometimes required in order to isolate pure recombinant vectors.

Two alternative procedures in which no parental infectious viral genome is used have been described (5, 24, 29). One method is based on the use of overlapping Ad5 DNA sequences cloned in two bacterial plasmids (5). The first plasmid carries the total Ad5 genome with a deletion of the DNA packaging signal and part of the E1 region, with the left and right inverted terminal repeat (ITR) sequences directly covalently joined. The second plasmid contains the left-end viral sequences, including the packaging signal and a passenger gene in place of the E1 region. Cotransfection of these plasmids in 293 cells results in the production of infectious recombinant vectors by in vivo homologous recombination. The configuration of the first plasmid is however known to be unstable in *Escherichia coli* (13, 28, 31). Moreover, the introduction of specific mutations or deletions in regions other than E1 requires tedious preliminary cloning steps in *E. coli*. The second alternative method is based on the manipulation of the full-length Ad genome as an infectious yeast artificial chromosome (YAC) (24). Targeted modifications of the viral genome are introduced by homologous recombination in yeast cells, and infectious virions are generated after transfection of the adenovirus genome, excised from the YAC vector, into appropriate cells. Although powerful, this method requires the use of an additional host (yeast) in which DNA yields are relatively low.

In the present study, we describe a novel procedure for the generation of recombinant adenovirus vectors that takes advantage of the highly efficient homologous recombination machinery of *E. coli* (6, 10, 11, 25, 30). We applied this method to clone the full-length Ad5 genome in one unique and stable infectious bacterial plasmid and to modify specifically several viral genetic regions. The production of vectors constitutively expressing the E3 region and/or bearing a human coagulation factor IX (hFIX) expression cassette in place of E1 is described in this paper as an example of the flexibility of this method.

The full-length 36-kb Ad5 genome was first cloned in a 2-kb plasmid by homologous recombination in *E. coli* between Ad5 virion DNA and a linear form of the pTG3601 plasmid (Fig. 1A; Table 1). pTG3601 was derived from the ppolyII plasmid (27) by insertion of 935- and 853-bp Ad5 DNA fragments amplified by PCR from the left and right ends of the Ad5 genome, respectively. Cotransformation of the *Bgl*II-linearized pTG3601 plasmid and Ad5 genomic DNA into *E. coli* BJ5183 *recBC sbcBC* (18) regenerated a stable circular pTG3602 plasmid containing the total Ad5 genome and conferring ampicillin resistance to the bacterium (Fig. 1). The frequency of recombinants was high, as estimated by the ratio between the number of ampicillin-resistant colonies obtained after cotransformation of *Bgl*II-digested pTG3601 and Ad5 DNA and the number of colonies obtained after transformation of the linearized pTG3601 plasmid alone. Depending on the experimental conditions, this ratio was between 5 and 61 (Table 1). This ratio indicates that the linearization of the plasmid prevents the growth of background colonies containing the parental vector. The efficiency of recombination could, however, not be significantly improved by altering the insert (I):vector (V) ratio (Table 1). One explanation might be that the size of the Ad5

* Corresponding author. Mailing address: Transgène S.A., 11 rue de Molsheim, 67000 Strasbourg, France. Phone: 33 88 27 91 00. Fax: 33 88 27 91 11.



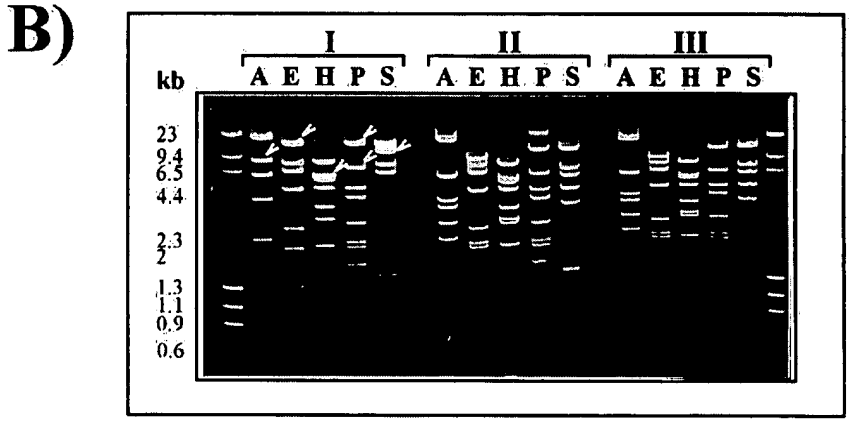


FIG. 1. Recombinational cloning of infectious full-length Ad5 genome in *E. coli*. (A) *E. coli* BJ5183 was cotransformed with Ad5 genomic DNA and a *Bgl*II-linearized pTG3601 plasmid containing 935 and 853 bp from the left and right ends of Ad5, respectively. In vivo homologous recombination generates ampicillin-resistant colonies containing the 38-kb pTG3602 plasmid carrying the full-length Ad5 genome. Large-scale preparation of pTG3602 DNA was carried out in the C600 *E. coli* strain. Cloned Ad5 DNA is still fully infectious as shown by the development of virus plaques after transfection of permissive human cells with *Pac*I-restricted pTG3602 DNA. (B) Restriction endonuclease analysis of pTG3602 plasmid DNA (lanes I), DNA purified from AdTG3602 virus produced by transfection of *Pac*I-restricted pTG3602 DNA (lanes II), and wild-type Ad5 DNA (lanes III). Plasmid DNA (1 μg), 10 μl of viral DNA prepared according to the method of Hirt (20), or 1 μg of purified Ad5 DNA was digested with *Afl*II (A), *Eco*RV (E), *Hin*dIII (H), *Pvu*I (P), and *Sph*I (S). Carets mark the positions of bands corresponding to the plasmid-Ad5 junctions. Size markers are indicated on the left.

genome (36 kb) constitutes a limiting factor for the efficiency of transformation. This possibility is supported by the observation that *E. coli* transformation is inhibited when increased amounts of Ad5 DNA are used (Table 1).

Eight bacterial colonies randomly isolated for further characterization were all shown by DNA restriction analysis to contain the expected plasmid (Table 1), suggesting that each recombination event occurred through homologous pairing of free viral DNA ends. However, plasmid DNA yields were low in BJ5183 cells, probably because of the formation of plasmid multimers (26). pTG3602 DNA was therefore retransformed into the C600 bacterial strain (22) to isolate larger quantities of plasmid DNA. Extensive characterization of pTG3602 DNA by restriction endonuclease analysis failed to detect any visible DNA rearrangement (Fig. 1B). The stability of the 38-kb plasmid is presumably due to the insertion of the 2-kb ppolyII sequence between the two viral ITRs, avoiding the constitution of palindromic sequences (28, 31).

Two unique *Pac*I sites were introduced by PCR immediately upstream of the left ITR and downstream of the right ITR in the initial pTG3601 plasmid (Fig. 1A). Since *Pac*I is absent in Ad5 genomic DNA, *Pac*I digestion allows the precise excision

TABLE 1. Efficiency of recombinational cloning of full-length Ad5 genome

Vector pTG3601-*BglII* (ng)	Insert Ad5 DNA (ng)	I:V molar ratio[a]	No. of Amp^r colonies	HR/V[b]
1	0		4	
	9.4	1	245	61
	94	10	136	34
10	0		64	
	94	1	627	10
	940	10	321	5

[a] I, insert; V, vector.
[b] The efficiency of homologous recombination (HR) is indicated by the HR/V ratio, calculated as the ratio of the number of colonies generated by the cotransfected vector and Ad5 insert DNA and the number of colonies generated by the vector alone.

of the full-length Ad5 genome from the pTG3602 plasmid. The infectivity of this pTG3602-derived Ad5 genome was demonstrated by calcium phosphate transfection (17) of 293 and A549 cells (34): large numbers of plaques were observed in cells transfected with the *Pac*I-restricted pTG3602 DNA, while the closed circular plasmid was unable to generate any viral plaques, confirming that at least one ITR extremity has to be in a free configuration to allow efficient adenovirus DNA replication (Table 2) (4, 19). These results show that the plasmid-derived Ad5 genome is fully infectious, with a specific infectivity of about 1/20 to 1/40 of that of Ad5 genomic DNA purified from wild-type virus particles (Table 2). Progeny virus recovered from independent pTG3602 plaques was amplified on 293 cells and further analyzed for growth characteristics, virus production yields, and DNA restriction patterns. In all cases, pTG3602-derived adenovirus (AdTG3602) was indistinguishable from wild-type Ad5 (Fig. 1B).

TABLE 2. Specific infectivity of cloned wild-type and recombinant Ad5 genomes[a]

DNA	DNA concn	No. of plaques/6-cm dish	
		293 cells	A549 cells
Ad5	0.1	>50, >50	4, 5
	1	Lysed, lysed	23, 34
pTG3602	1	0, 0	0, 0
	5	0, 0	0, 0
pTG3602-*Pac*I	1	23, 33	1, 0
	5	Lysed, lysed	6, 10
pTG3604-*Pac*I	1	53	1
	5	Lysed	13
pTG3614-*Pac*I	1	20, 41	ND
	5	Lysed, lysed	ND
pTG3606-*Pac*I	1	52, 27	ND
	5	Lysed, lysed	ND
pTG3622-*Pac*I	1	17, 9	ND
	5	63, 60	ND
pTG3623-*Pac*I	1	17, 17	ND
	5	34, 22	ND

[a] A549 and/or 293 cells were transfected with 0.1 and 1 μg of Ad5 DNA or nonrestricted or *Pac*I-restricted plasmid DNA. Cells were overlaid with agar 15 h later, and plaques were then counted 14 days posttransfection. ND, not done (these viruses have E1 deletions and do not grow on A549 cells). Results from two independent experiments are shown (except for those for pTG3604-*Pac*I).

Since construction of recombinant adenovirus vectors is usually a time-consuming procedure, we then tested whether a single-step replacement strategy exploiting the *E. coli* homologous recombination machinery could be designed to modify selectively any particular genetic region of Ad5 (Fig. 2). The viral region to be modified is first subcloned into a bacterial plasmid and the desired deletions, insertions, or mutations are performed by conventional molecular biology techniques (32). The modified segment is then purified and cotransformed into BJ5183 together with the appropriately restricted plasmid containing the full-length Ad5 genome (Fig. 2). Homologous recombination events between the free ends of the modified replacement fragment and their homologous viral sequences in the linearized vector eventually generate plasmids containing Ad5 genomes with the required modifications.

As an example, we derived from pTG3602 a new plasmid (pTG3604) containing a full-length Ad5 genome in which expression of E3 (36) was relieved from the control by E1 proteins and made constitutive by the introduction of the Rous sarcoma virus 3′ long terminal repeat promoter sequences (LTR_{RSV}). LTR_{RSV} was inserted in front of E3 at nucleotide (nt) 28249 (throughout the manuscript, nucleotide numbers refer to positions on the Ad5 genome, according to reference 8) in a transfer plasmid (pTG1696) bearing an Ad5 segment from nt 21562 to the right-end ITR. This insertion site was chosen in order to maintain the integrity of the L4 mRNA polyadenylation signal which overlaps the E3 mRNA transcription initiation site (33). A *BglI*$^{(nt\ 24862–35211)}$ fragment encompassing the modified E3 region was then isolated from pTG1696 and recombined in *E. coli* with pTG3602 previously linearized at nt position 27082 by *Spe*I digestion (Fig. 2A). *Spe*I cuts the Ad5 genome upstream of the targeted E3 region and leaves three regions of DNA homology with the pTG1696 *BglI* fragment: (i) 2,220 bp between the *BglI*$^{(nt\ 24862)}$ and *SpeI*$^{(nt\ 27082)}$ sites, (ii) 1,167 bp between *SpeI*$^{(nt\ 27082)}$ and the 5′ boundary (nt 28249) of LTR_{RSV}, and (iii) 7,623 bp between the 3′ boundary (nt 27588) of LTR_{RSV} and the *BglI*$^{(nt\ 35211)}$ site. As a consequence, homologous recombination should generate two types of plasmids (Fig. 2A): (i) the parental pTG3602 plasmid reconstituted by recombination between sequences upstream from the *Spe*I site and sequences located between *Spe*I and the E3 region and (ii) the expected pTG3604 plasmid containing the E3-modified genome produced by recombination between sequences upstream from *Spe*I and downstream from the E3 region. The radioactive screening of the ampicillin-resistant colonies with an LTR_{RSV} oligonucleotide probe confirmed this hypothesis since 18.5% of the 270 colonies were found to contain the pTG3604 plasmid (Table 3). DNA analysis performed on six randomly selected candidates confirmed that the cloned viral genome was genetically stable and full-length and contained a modified E3 region. Moreover, homologous recombination in BJ5183 was highly efficient since 270 ampicillin-resistant colonies were obtained after cotransformation with the linearized pTG3602 plasmid and the pTG1696 fragment, while only 2 colonies appeared with the linearized pTG3602 vector alone (Table 3).

Similar to E3, the E1 region can also be efficiently modified by homologous recombination in *E. coli*. As an example (Fig. 2B and C), the E1 region of Ad5 was either deleted or replaced by an expression cassette encoding hFIX (2). Replacement of E1 by hFIX was done by cotransformation of *E. coli* BJ5183 with a pTG3602 or pTG3604 plasmid linearized in E1 by *ClaI*$^{(nt\ 918)}$ digestion and an *Msc*I DNA fragment bearing the hFIX cDNA under the control of the mouse phosphoglycerate kinase gene

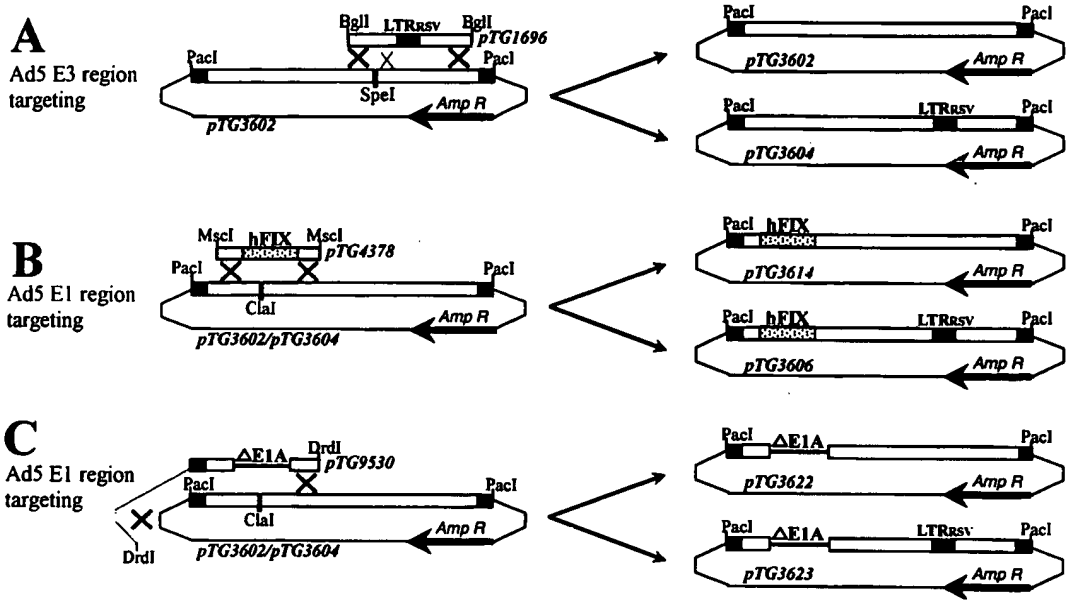


FIG. 2. Construction of recombinant adenovirus vectors by homologous recombination in *E. coli*. Several strategies of Ad5 genome manipulation are presented: insertion of a new regulatory sequence (LTR_{RSV} promoter) upstream of the Ad5 E3 region (A), replacement of the Ad5 E1 region by a foreign sequence (expression cassette for human coagulation factor IX) (B), and deletion of the Ad5 E1 region (C). ×, homologous recombination events required for the targeted modification of the viral DNA; ×, an homologous recombination event that reconstitutes the parental target vector (panel A). ITRs (closed boxes), Ad5 genomic DNA (open boxes), foreign sequences (stippled boxes), E1 deletion (thick line), plasmid sequences (thin lines), β-lactamase gene (arrows), and plasmid excision (labeled *PacI*) and linearization (labeled *BglII*, *ClaI*, and *SpeI*) restriction sites are also indicated.

promoter (1) and the simian virus 40 polyadenylation signal (Fig. 2B). The *MscI* insert was isolated from a transfer plasmid (pTG4378) containing the hFIX expression cassette in place of the E1 region (nt 459 to 3328), in an adenovirus DNA fragment corresponding to 17% of the left genomic sequence (nt 1 to 6241). Excision by *MscI* leaves upstream and downstream from the hFIX cassette 188 and 2,044 bp, respectively, of Ad5 sequences homologous to viral DNA in pTG3602 and pTG3604 (Fig. 2B). Cotransformation into BJ5183 generated only few ampicillin-resistant colonies (Table 3), probably as a consequence of the short length (188 bp) of sequence with DNA homology upstream of the *ClaI* site (see below). However, linearization of target vectors in the E1 region significantly increased the frequency of colonies containing plasmids in which E1 was replaced by the hFIX expression cassette: 1 positive colony of 2 and 4 positive colonies of 17 identified for the pTG3602- and pTG3604-derived vectors, respectively (Table 3). The negative colonies contained the background pTG3602 and pTG3604 parental vectors, probably corresponding to residual uncut vectors.

TABLE 3. Efficiency of targeted modifications of the cloned Ad5 genome[a]

Insert	Vector	HR/V[b]	No. of positive colonies/total no. of colonies (plasmid)
pTG1696-*BglI*	pTG3602-*SpeI*	270/2	50/270 (pTG3604)
pTG4378-*MscI*	pTG3602-*ClaI*	2/0	1/2 (pTG3614)
	pTG3604-*ClaI*	17/19	4/17 (pTG3606)
pTG9530-*DrdI*	pTG3602-*ClaI*	101/1	6/6 (pTG3622)
	pTG3604-*ClaI*	569/48	6/6 (pTG3623)

[a] Plasmids carrying wild-type (pTG3602) or E3-modified (AdTG3604) Ad5 genomes were digested by *SpeI* for E3 region targeting or by *ClaI* for E1 region targeting and transformed into *E. coli* BJ5183 either alone or together with a 10-fold molar excess of the purified replacement insert. Positive colonies containing plasmids with the modified full-length Ad5 genomes were identified by radioactive screening or with a plasmid minipreparation.

[b] Number of colonies generated by the cotransfected vector and insert per the number of colonies generated by the vector alone.

We tested whether increasing the length of the region with DNA homology upstream of the E1-replacement cassette might improve the homologous recombination efficiency (30). We used a transfer plasmid (pTG9530) containing an Ad5 left-end region (nt 1 to 5788) deleted from part of the E1 sequences (nt 459 to 3328) in the same ppolyII plasmid backbone as pTG3602. A *DrdI* fragment isolated from pTG9530 was cointroduced into bacteria along with the *ClaI*-restricted pTG3602 or pTG3604 target vector (Fig. 2C). The sequence of DNA homology upstream of the *ClaI* site was 2,244 bp long (compared with 188 bp in the previous experiment), and this increase in length dramatically affected the HR/V ratio, which increased to 101/1 and 569/48 for the pTG3602- and pTG3604-derived vectors, respectively (Table 3). Analysis of six randomly picked colonies showed the presence of the expected pTG3622 (full-length Ad5 deleted from E1) and pTG3623 (full-length Ad5 deleted from E1 and with a constitutive E3 region) recombinant plasmids.

These experiments demonstrate that the replacement or the deletion of regulatory or structural viral genes can be easily achieved by homologous recombination in *E. coli*. The advantages of using *E. coli* instead of yeast cells (24) include higher transfection efficiency, higher growth rates, and higher plasmid yields. Moreover, the *E. coli* single-step replacement strategy is technically more straightforward than the two-steps replacement strategy classically used to target a segment of yeast DNA. We observed that the frequency of homologous recombination events and the efficiency of recovery of the expected recombinants could vary depending on the design of the DNA

transfer strategy. In the described examples (Fig. 2B and C; Table 3), the frequency of homologous recombination could be enhanced 10- to 100-fold by increasing the length of the region with DNA homology. Moreover, the yield of positive recombinants was higher when the vector was linearized precisely in the targeted region, as in the E1 replacement experiments (Fig. 2B and C). Modification of the E3 region was done with a target vector linearized at an *Spe*I site located 1,167 bp upstream from E3. As a consequence, two types of plasmids were obtained: the parental pTG3602 vector and the E3-modified Ad5 genome (Fig. 2A). The same phenomenon was observed in deletion experiments targeting the fiber gene or the E4 region which is located 4.8 or 5.8 kb downstream of the *Spe*I site, respectively: the frequency of correct recombination events was found to be inversely proportional to the distance between the targeted region and the *Spe*I site (unpublished observations). The percentage of positive colonies was nevertheless always >1%, and the identification of these colonies by classical screening techniques was straightforward.

Transfection of *Pac*I-restricted pTG3604, pTG3606, pTG3614, pTG3622, and pTG3623 plasmid DNA into 293 cells allowed the production of viral plaques with efficiencies similar to that obtained with pTG3602 (Table 2). In all cases, transfection of as little as 1 μg of viral DNA was sufficient to generate enough viral plaques for virus propagation and purification; the specific infectivity was around 30 plaques per μg of linearized adenovirus DNA in 293 cells (Table 2). As a comparison, adenovirus DNA excised from a YAC was shown to generate 2 to 10 plaques per μg of total yeast DNA, corresponding to a calculated specific infectivity of 100 to 500 plaques per μg of viral DNA (24). The method recently described by Bett et al. (5) and McGrory et al. (29) generates 4 to 8 plaques per μg of transfected plasmid DNA. In the latter case, the transfer plasmid is cotransfected with a circular plasmid bearing the Ad5 genome, which is in a relatively unstable configuration and is less infectious than linear DNA (12, 13). The procedures described here and by Ketner et al. (24) are based on prior manipulation of the full-length viral genome contained in single molecules of stable YACs or plasmids. Transfection into permissive eukaryotic cells of Ad5 genome excised from YAC or plasmid sequences allows the systematic recovery of pure virus plaques. Identification of the positive recombinant viruses by plaque screening and isolation of pure virus stocks by repeated plaque purifications are therefore not required for these methods. As a consequence, the construction of recombinant vectors by the described procedure is usually faster than that by previous methods; as an example, production of pure plaques of a vector in which E1 is deleted with pTG3602 and the E1-transfer plasmid as starting materials requires at most 20 days compared with at least 6 weeks for methods in which similar starting materials (Ad5 DNA and E1-transfer plasmid) are cotransfected in 293 cells (3, 14).

The analysis of the phenotype of all plasmid-derived viruses confirmed that the selected modifications were correctly and stably introduced into the viral genomes. As an illustration, we have shown that expression of E3 in AdTG3606 is no longer controlled by E1 but is driven by the constitutive LTR_{RSV} promoter sequences inserted upstream from the E3 coding region (Fig. 3A). 293 and A549 cells were infected with wild-type Ad5 or E1-deleted AdTG3606 or AdTG3614 viruses at multiplicities of infection of 10 (data not shown) and 100 (Fig. 3A). Radioimmunoprecipitation of the cell extracts with monoclonal antibody Tw1.3 (9) directed against the E3-encoded gp19K protein showed an expression of the viral glycoprotein in both cell lines infected with either Ad5 or AdTG3606. In contrast, infection with a virus with an E1 deletion and carrying wild-type E3 sequences (AdTG3614) led to gp19K expression only in the E1-expressing 293 cells (Fig. 3A). In addition, the intravenous administration of AdTG3606 or AdTG3614 containing an hFIX expression cassette in place of E1 led to high concentrations of hFIX in C57BL/6 mice: 3 days after virus injection, mean levels of 300 and 650 ng/ml were detected in the sera of the animals treated with AdTG3606 and AdTG3614, respectively (Fig. 3B).

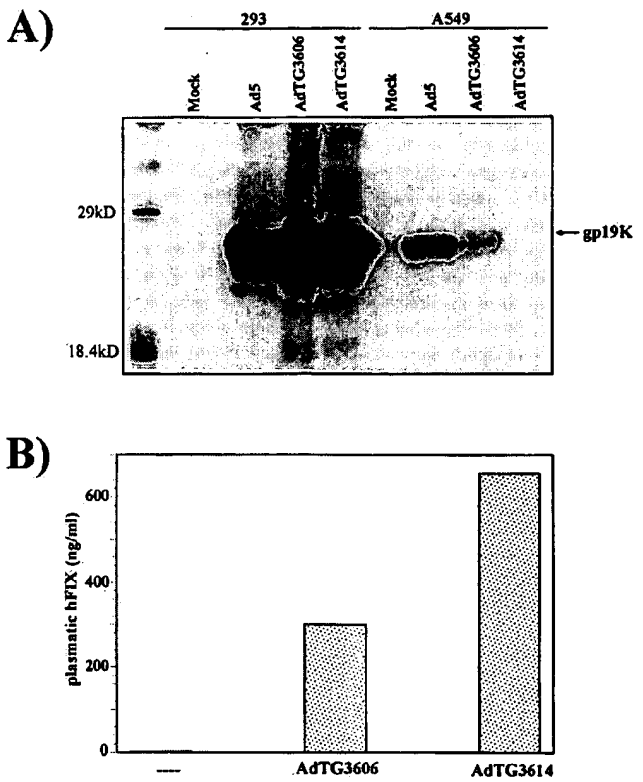


FIG. 3. Foreign sequences introduced into recombinant viruses by in vivo homologous recombination are functional. (A) Insertion of the LTR_{RSV} promoter sequences upstream of the Ad5 E3 region confers constitutive expression to gp19K. 293 and A549 cells were mock infected or infected for 24 h with AdTG3606, AdTG3614, or wild-type Ad5. Total proteins labeled with [^{35}S]methionine/cysteine were immunoprecipitated with MAb Tw1.3, an anti-gp19K monoclonal antibody, and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. (B) purified AdTG3606 or AdTG3614 recombinant vector in which the E1 region was replaced by an expression cassette for hFIX was injected into the tail vein of three C57BL/6 mice. Serum was recovered at 3 days postinjection and analyzed by enzyme-linked immunosorbent assay for the plasmatic concentration in human factor IX. The means of data for each set of animals is shown. This first lane (– –) shows the results for a control, untreated mouse. In panel A, size markers (in kilodaltons [kD]) are indicated on the left.

The procedure described in this report allows the rapid and efficient cloning and manipulation of full-length infectious adenovirus genomes in bacterial plasmids. This method combines the powerful genetic engineering techniques that are available in *E. coli* and the ability of this microorganism to recombine homologous sequences at a high frequency (6, 10, 11, 25, 30). The advantages of this technology are multiple: (i) all cloning and, more importantly, recombination steps are carried out in *E. coli*, (ii) the frequency of bacterial colonies containing the plasmid with the modified adenovirus genome is very high, (iii) any genetic region of the viral genome can be specifically modified or deleted if appropriate restriction sites are available, (iv) plasmids containing the full-length adenovirus genomes can be introduced into appropriate bacterial strains for production of large amounts of viral DNA, and (v) transfection

of excised recombinant adenovirus DNA into permissive human cells generates plaques containing only pure virus particles. We demonstrate that full-length recombinant adenovirus genomes modified in both E1 and E3 regions can be efficiently generated and that their rescue as pure viral particles was guaranteed in transfected 293 cells. We similarly produced adenoviruses with either deletions of or modifications in the fiber gene, the E2A gene, or the E4 region (unpublished data). The production of even-further-crippled viral vectors is theoretically possible with the same technology, provided that adequate restriction sites are available for Ad5 DNA linearization. The presence of the *Cla*I, *Bam*HI, and *Spe*I sites located at positions 918, 21562, and 27082, respectively, already allows efficient modifications of most of the viral genes. However, improvement of this approach is possible by the introduction of new unique sites at other locations. Investigation of these sites is currently in progress.

We thank J. W. Yewdell for kindly providing the MAb Tw1.3 anti-gp19K monoclonal antibody. We also thank H. Schultz, D. Ali-Hadji, P. Kopp, and I. Mauclair for excellent technical assistance and M. Courtney and M. Lusky for critical reading of the manuscript.

This work was supported in part by the Association Française contre la Mucoviscidose and the Association Française contre les Myopathies.

REFERENCES

1. **Adra, C. N., P. H. Boer, and M. W. McBurney.** 1987. Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. Gene **60**:65–74.
2. **Balland, A., T. Faure, D. Carvallo, P. Cordier, P. Ulrich, B. Fournet, H. DeLaSalle, and J. P. Lecocq.** 1988. Characterization of two differently processed forms of human recombinant factor IX synthesized in CHO cells transformed with a polycystronic vector. Eur. J. Biochem. **172**:565–572.
3. **Berkner, K. L.** 1988. Development of adenovirus vectors for the expression of heterologous genes. BioTechniques **6**:616–629.
4. **Berkner, K. L., and P. A. Sharp.** 1983. Generation of adenovirus by transfection of plasmids. Nucleic Acids Res. **11**:6003–6020.
5. **Bett, A. J., W. Haddara, L. Prevec, and F. L. Graham.** 1994. An efficient and flexible system for construction of adenovirus vectors with insertions or deletions in early regions 1 and 3. Proc. Natl. Acad. Sci. USA **91**:8802–8806.
6. **Bubeck, P., M. Winkler, and W. Bautsch.** 1993. Rapid cloning by homologous recombination *in vivo*. Nucleic Acids Res. **21**:3601–3602.
7. **Chengalvala, M. V. R., M. D. Lubeck, B. J. Selling, R. J. Natuk, K.-H. L. Hsu, B. B. Mason, P. K. Chanda, R. A. Bhat, B. M. Bhat, S. Mizutani, A. R. Davis, and P. P. Hung.** 1991. Adenovirus vectors for gene expression. Curr. Opin. Biotechnol. **2**:718–722.
8. **Chroboczek, J., F. Bieber, and B. Jacrot.** 1992. The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. Virology **186**:280–285.
9. **Cox, J. H., J. R. Bennink, and J. W. Yewdell.** 1991. Retention of adenovirus E19 glycoprotein in the endoplasmic reticulum is essential to its ability to block antigen presentation. J. Exp. Med. **174**:1629–1637.
10. **Degryse, E.** 1995. Evaluation of *Escherichia coli recBC sbcBC* mutants for cloning by recombination *in vivo*. J. Biotechnol. **39**:181–187.
11. **Degryse, E.** 1996. *In vivo* intermolecular recombination in *Escherichia coli*: application to plasmid constructions. Gene, **170**:45–50.
12. **Fisher, K. J., H. Choi, J. Burda, S. J. Chen, and J. M. Wilson.** 1996. Recombinant adenovirus deleted of all viral genes for gene therapy of cystic fibrosis. Virology **217**:11–22.
13. **Ghosh-Choudhury, G., Y. Haj-Ahmad, P. Brinkley, J. Rudy, and F. L. Graham.** 1986. Human adenovirus cloning vectors based on infectious bacterial plasmids. Gene **50**:161–171.
14. **Graham, F. L., and L. Prevec.** 1991. Manipulation of adenovirus vectors. Methods Mol. Biol. **7**:109–128.
15. **Graham, F. L., and L. Prevec.** 1992. Adenovirus based expression vectors and recombinant vaccines, p. 363–390. *In* R. W. Ellis (ed.), Vaccines: new approaches to immunological problems. Butterworths-Heineman, London.
16. **Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J. Gen. Virol. **36**:59–72.
17. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
18. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557–580.
19. **Hanahan, D., and Y. Gluzman.** 1984. Rescue of functional replication origins from embedded configurations in a plasmid carrying the adenovirus genome. Mol. Cell. Biol. **4**:302–309.
20. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
21. **Horwitz, M. S.** 1990. Adenoviridae and their replication, p. 1679–1721. *In* B. N. Fields, D. M. Knipe, et al. (ed.), Virology, 2nd ed. Raven Press, Ltd., New York.
22. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening of cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning. IRL Press, Oxford.
23. **Imler, J.-L., C. Chartier, A. Dieterlé, D. Dreyer, M. Mehtali, and A. Pavirani.** 1995. An efficient procedure to select and recover recombinant adenovirus vectors. Gene Ther. **2**:263–268.
24. **Ketner, G., F. Spencer, S. Tugendreich, C. Connelly, and P. Hieter.** 1994. Efficient manipulation of the human adenovirus genome as an infectious yeast artificial chromosome clone. Proc. Natl. Acad. Sci. USA **91**:6186–6190.
25. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer.** 1994. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. **58**:401–465.
26. **Kusano, K., K. Nakayama, and H. Nakayama.** 1989. Plasmid-mediated lethality and plasmid multimer formation in *Escherichia coli recBC sbcBC* mutant. J. Mol. Biol. **209**:623–634.
27. **Lathe, R., J. L. Vilotte, and A. J. Clark.** 1987. Plasmid and bacteriophage vectors for excision of intact inserts. Gene **57**:193–201.
28. **Leach, D. R. F., and F. W. Stahl.** 1983. Viability of λ phages carrying a perfect palindrome in the absence of recombination nucleases. Nature (London) **305**:448–451.
29. **McGrory, W. J., D. S. Bautista, and F. L. Graham.** 1988. A simple technique for the rescue of early region I mutations into infectious human adenovirus type 5. Virology **163**:614–617.
30. **Oliner, J. D., K. W. Kinzler, and B. Vogelstein.** 1993. *In vivo* cloning of PCR products in *E. coli*. Nucleic Acids Res. **21**:5192–5197.
31. **Raleigh, E., K. Lech, and R. Brent.** 1994. Selected topics from classical bacterial genetics, p. 1.4.1–1.4.14. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology, vol. 1, suppl. 11.
32. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Sittler, A., H. Gallinaro, and M. Jacob.** 1994. Upstream and downstream *cis*-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2. Nucleic Acids Res. **22**:222–231.
34. **Smith, C. D., D. W. Craft, R. S. Shiromoto, and P. O. Yan.** 1986. Alternative cell line for virus isolation. J. Clin. Microbiol. **24**:265–268.
35. **Trapnell, B. C.** 1993. Adenoviral vectors for gene transfer. Adv. Drug Deliv. Rev. **12**:185–199.
36. **Wold, W. S. M., and L. R. Gooding.** 1991. Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus-cell interactions. Virology **184**:1–8.

ELSEVIER

Mutation Research 400 (1998) 45–58

MMR
Fundamental and Molecular Mechanisms of Mutagenesis

Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability

D.A. Gordenin *,[1], M.A. Resnick

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, 101 Alexander Dr., P.O. Box 12233, Research Triangle Park, NC 27709, USA

Received 22 January 1998; revised 18 February 1998; accepted 20 February 1998

Abstract

The genomes of all organisms contain an abundance of DNA repeats which are at-risk for causing genetic change. We have used the yeast *Saccharomyces cerevisiae* to investigate various repeat categories in order to understand their potential for causing genomic instability and the role of DNA metabolism factors. Several types of repeats can increase enormously the likelihood of genetic changes such as mutation or recombination when present either in wild type or mutants defective in replication or repair. Specifically, we have investigated inverted repeats, homonucleotide runs, and short distant repeats and the consequences of various DNA metabolism mutants. Because the at-risk motifs (ARMs) that we characterized are sensitive indicators, we have found that they are useful tools to reveal new genetic factors affecting genome stability as well as to distinguish subtle differences between alleles.

Keywords: Mutation; Recombination; Repeat; Genome stability; Mutator

1. Introduction

The balance between genome integrity and genetic instability is determined by a vast number of factors, most of which have been addressed at one time or another in the previous 399 volumes of this Journal. In recent years, it has become apparent that repeated DNA sequences can have a high potential for mutation and recombination which is greatly influenced by repair or replication capabilities. Since the likelihood of repeat induced genetic change usually depends on arrangement rather than specific sequence, we have referred to these motifs as at-risk motifs or ARMs [1]. Motifs that can lead to non-canonical DNA structures [2] are a particularly interesting category of ARMs. For example, they might lead to novel structures during replication that are refractory to error correction or they might stimulate further chromosomal changes such as recombination.

We have used the yeast *Saccharomyces cerevisiae* to identify a variety of ARMs that can give rise to non-canonical DNA structures and we have investigated the genetic controls of ARM-associated genomic instability. Because of their high potential for change, ARMs represent important threats to genome stability. Since these ARMs have human

Abbreviations: ARMs, at-risk motifs; MMR, mismatch repair; DNA-Pol, DNA polymerase; DSB, double-strand break; IR, inverted repeat

* Corresponding author. Tel.: +1-919-541-5190; fax: +1-919-541-7593; E-mail: gordenin@niehs.nih.gov

[1] On leave from the Department of Genetics, St. Petersburg State University, St. Petersburg 199034, Russian Federation.

0027-5107/98/$19.00
PII S0027-5107(98)00047-5

counterparts and there are many examples of human DNA metabolic gene homologues, the systems are also expected to provide an understanding of threats from ARMs in humans.

2. Inverted repeats

Large direct repeats have long been known to be capable of undergoing recombination. However, inverted repeats (IRs) exhibit novel genome destabilizing features, some of which were initially discovered by Collins [3] and Collins et al. [4]. IRs were the first members of an ARMs category whose effects are attributed to the generation of non-canonical DNA structures. Collins et al. showed that head-to-head IRs (palindromes) were deleted at extremely high rates in *Escherichia coli*. Palindromes comprised of at least 485 bp repeats were unable to propagate in bacterial cells; only partial or complete deletions could be recovered. The proposed mechanism of IR-stimulated deletions includes the generation of a hairpin structure in single strand DNA (ssDNA) that arise during replication as described in Fig. 1 (for reviews, see Refs. [5,6]). The non-canonical DNA hairpin structure would impair elongation of the nascent strand because the DNA polymerase (DNA-Pol) is stalled at the base of the hairpin.

Thus, the IR group of ARMs appear to exert their effect by forming structures that interfere with normal DNA metabolism. The observed deletions could arise by several processes. The hairpin could be cleaved by a structure-specific nuclease and the surrounding DNA could be end-joined [5–8]. Alternatively, the hairpin could lead to replication slippage between fortuitous short repeats that might be present near the base of the stem as shown in Fig. 1 [3,9–11].

Subsequent studies based in yeast indicate that IRs can stimulate deletions in eukaryotes via replication mechanisms. In these studies, which took advantage of the replicative DNA polymerases (DNA-Pol) α and δ mutants, the deletion likelihood was increased in the replication defective strains [12–14]. The deletion rates of various IRs, such as the 1.5 kb IRs of bacterial transposon *Tn5* separated by a long (2.7 kb) spacer, short palindromes or quasipalindromes (IRs separated by only 8–9 bp), were increased in the DNA-Pol mutants. (Note: bacterial transposon *Tn5* was used in yeast studies [5,15] as a convenient DNA insert containing IRs flanked by short direct repeats (Fig. 1).) While the increase for unstable palindromes and quasipalindromes was moderate, a very strong effect was observed for the *Tn5* which is stable in wild type cells. Deletions of this IR were increased up to 1000-fold in DNA-Pol δ and α temperature sensitive mutants [5,15], demonstrating that stable ARMs can have a large potential for change.

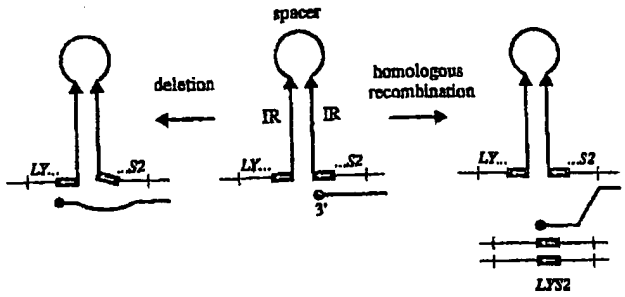


Fig. 1. Replication model for IR generated genomic rearrangements. This model is based on results with model IRs inserted in the *LYS2* gene of yeast [5–7]. IR-insert flanked by short direct repeats (arrows inside black boxes) inactivates the gene *LYS2*. During replication of the IR sequence (long arrows), single-stranded regions of DNA may give rise to a stem-like secondary structure, thereby inhibiting extension of the 3'-end and possibly leading to deletion or recombination. Deletion of the IR (precise excision leading to Lys^+ phenotype) could result from replication slippage between short direct repeats at the base of the secondary structure. Alternatively, the stalled replication could initiate recombination with a homologous *LYS2* sequence. (In our constructs, this *LYS2* sequence contains a mutation; thus, wild type recombinants can be selected based on their Lys^+ phenotype. Possible events initiating recombination are discussed in the text). If recombination involves identical sequences in sister chromatids, it will not result in rearrangements, while recombination within the same chromatid or another chromosome can result in deletions or translocations.

We found that an IR can stimulate not only deletions but also recombination in adjacent regions. IRs elevated interchromosomal allelic recombination [5], recombination between repeats in nonhomologous chromosomes [6] and intrachromosomal recombination between homologous or highly diverged repeats [6,16]. As expected, some of the IR-stimulated intrachromosomal and interchromosomal recombination events were associated with deletions and translocations. The presence of a strong bi-directional origin of replication in the spacer region within

an IR reduced the rates of deletion and recombination [6]. This supports the view that IR stimulated events are initiated by secondary structures formed between single stranded DNA regions of IRs during replication. It is interesting that high levels of intrachromosomal recombination have been found within IRs of a long (15 kb) palindrome sequence introduced into mice [10].

We proposed that arrest of the replication complex at the base of a hairpin formed by IRs in ssDNA could initiate IR-associated recombination as well as deletions [5–7]. Homologous recombination would, therefore, be an alternative to replication slippage when DNA elongation is blocked by a non-canonical DNA structure (Fig. 1). Three mechanisms have been proposed for IR stimulated recombination. (i) Recombination could be initiated by the subsequent formation of double-strand breaks (DSB), which have long been known to be highly recombinagenic [17]. It was suggested [18,19] and recently demonstrated in *E. coli* [20] that impaired DNA strand elongation can result in DSBs. Alternatively, DSBs could result from cleavage by structure specific endonucleases [9]. (ii) The stalled 3′ end of the nascent strand could separate from the template, forming a 3′-tail that could invade homologous double-stranded DNA (dsDNA). This is reminiscent of the model suggested for bypass of a DNA lesion by means of recombination-dependent replication in bacteriophage T4 [21]. (iii) A single-stranded region that can occur near the secondary structure formed by IRs in the template strand could initiate a search for homology [7] as a first step in recombination.

Both length of IRs and size of the spacer DNA between the repeats can influence deletion and recombination stimulated by this ARM [6]. The incidence of these genetic changes are directly related to the size of IRs and inversely related to the length of the spacer separating IRs. We found in this study that IRs are capable of stimulating exceptionally high levels of recombination leading to deletions and translocations in wild type yeast strains. A perfect palindrome formed by two head-to-head copies of a 1.0 kb *URA3* gene increased intra- and interchromosomal recombination in the adjacent region 2400-fold and 17,000-fold, respectively.

Thus, the IR category of ARMs can have a substantial effect on genome stability. Since mutations in a DNA replication polymerase appear to greatly increase IR associated changes, we suggest that such mutants represent a category of 'mutators' that exert their ARMs effects by increasing the likelihood of

Table 1
Sources of at-risk motif (ARM) instability and effects of mutators

ARM	Poor substrate for	Mutators that can destabilize ARM	Mutator mechanism
Inverted repeats	DNA-Pol	DNA-Pol	Premutational structures[a]
Homonucleotide runs	Proofreading	MMR	Lack of repair[b]
Short distant repeats (deletions)	MMR	DNA-Pol	Premutational structures
Short distant repeats (duplications)	Repair systems other than Rad27 (Fen1)	*FEN1 (RAD27)*	Lack of repair
Minisatellites	MMR (deletion)	DNA-Pol and *FEN1 (RAD27)*	Premutational structures
Triplet repeats	Fen1 (Rad27)	*(i) Alternative pathways to Fen1 (Rad27)*[c]; *(ii) changes causing more flaps*[c]	*(i) Lack of repair; (ii) Premutational structures*

The sources of mutator activity discussed in the text are named as: [a]Premutational structures—mutators increasing likelihood of non-canonical DNA structures that are poor substrates for DNA metabolism. [b]Lack of repair—mutators that inactivate a repair pathway that prevents a non-canonical DNA structure from becoming a mutation. [c]Proposed categories of mutators that could stimulate repeat expansion are shown by italics.

novel intermediates ('Premutation structure' in Table 1).

3. Homonucleotide runs and microsatellites

Microsatellites are multiply reiterated short (less than 14 bp as suggested by Sia et al. [22]) repeats. In humans, they are more variable than nonreiterated genomic sequences and the changes are usually due to small deletions or insertions of a small number of microsatellite units. Variable microsatellites have been useful as mapping markers in human genetics [23].

Microsatellites are extremely unstable in tumor cells deficient in post-replication mismatch repair (MMR) (for reviews, see Refs. [24–26]). MMR deficiency in humans was associated with an inherited predisposition to several types of cancer. It was proposed that the predisposition was due to an increased likelihood of inactivation of tumor suppressor genes in MMR deficient cells. Small microsatellites formed by repeats of one nucleotide (homonucleotide runs) in the coding sequences (CDS) of putative tumor suppressor genes appeared to be preferential sites of mutational inactivation in MMR deficient cells [27–29] suggesting that these might represent an important class of ARMs.

We have systematically examined the stability of homonucleotide run ARMs in wild type and MMR deficient yeast [30]. Extending the length of a homonucleotide sequence leads to an exponential increase in mutation rates. However, a MMR deficiency can greatly amplify the mutation rate. For example, the absence of MMR results in a 10,000-fold increased mutation rate in A_{13} and A_{14} homonucleotide runs. This is the largest impact of a single mutator observed in eukaryotes. For non-ARM DNA, strong mutator effects can be observed only when both DNA-Pol proofreading and MMR are inactivated [31,32]. The homonucleotide run ARM is more effective at increasing mutation rates caused by a MMR defect when compared with microsatellite ARMs containing longer (2–13 bp) repeat units [26]. This is probably due to a higher likelihood of replication slippage within the homonucleotide run as well as due to more efficient correction of smaller loops in wild type cells (see Ref. [33] and discussion below).

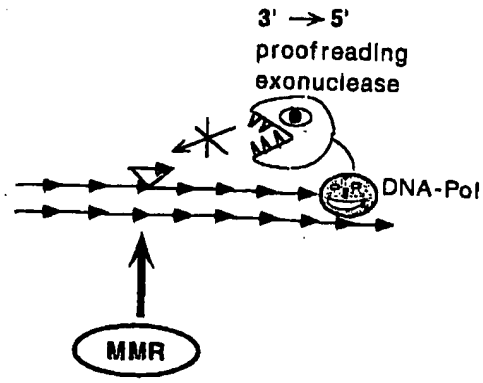


Fig. 2. DNA-Pol proofreading and mismatch repair preventing mutations caused by replication slippage within microsatellite DNA. Presented is a 'bulge' in the nascent strand that can lead to insertion of one repeat unit. Both DNA-Pol proofreading 3' – > 5' exonuclease and MMR can prevent this mutation. However, when a bulge is distant from the 3' end, it is proposed to be inaccessible to the proofreading exonuclease.

Based on results with *E. coli* and yeast [30,34] and with in vitro replication systems [35], microsatellites are hypermutable in MMR deficient cells because frameshift intermediates escape DNA-Pol proofreading (Fig. 2). Since proofreading and MMR act in series to remove replication errors [31,36], escape from DNA-Pol proofreading would leave the entire burden of preventing mutations within these ARMs to the MMR system.

While a MMR defect greatly enhances the potential of a homonucleotide run ARM to cause mutation, it differs from DNA-Pol defects on IRs which change the incidence of non-canonical DNA structure formed by ARMs. The MMR defect does not increase the number of mutation intermediates (bulges) in homonucleotide runs. Instead, the impact of the mutator is due to a lack of repair (Table 1) of premutational structures since mutation intermediates that arise in homonucleotide run ARMs are poor substrate for proofreading.

4. Short distant repeats

Short (< 10 bp) direct repeats separated by less than 100 bp can lead to duplications and extended deletions of the DNA between repeats (Fig. 3) in various prokaryotes and eukaryotes, especially in

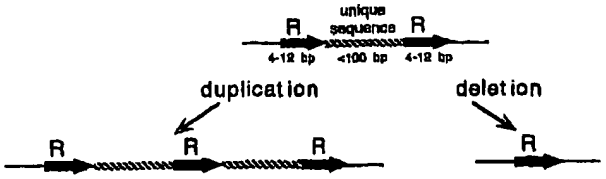


Fig. 3. Deletions and duplications associated with short direct repeats. Unique sequence that is either deleted or duplicated is indicated by the hatched line. The duplications and deletions involve gain or loss of one of the short direct repeats (R). The sizes of repeats and distances between repeats are based on studies in yeast [33,37,38].

cells with mutator mutations [8,33,37–40]. These changes most likely occur via end-joining repair of a double-strand break or via replication slippage between short repeats. Since the short distant repeats are frequent in all genomes (see below), they are an important category of ARMs that can cause rearrangements.

We developed a novel system that enabled us to study 31 bp and 61 bp deletions in the yeast *LYS2* gene (Fig. 4 and Refs. [33,37]). Deletions occurred between 6 bp or 7 bp short repeats flanking the insert that inactivates the *LYS2*. Deletions of the insert restore the *LYS2* function and can be selected as Lys^+ revertants. A role for replication slippage was supported by the observation that the spectrum of deletion breakpoints depended on orientation relative to replication origin. A temperature sensitive mutation in DNA-Pol δ (*pol3-t*) increased deletion rates up to 1000 fold, suggesting a role for replication in increasing the likelihood of genetic changes in these ARMs. (The *pol3-t* mutation is due to a G to A substitution resulting in a Gly_{641} to Ala_{641} change near the conserved region VI [7]. The mutation probably alters a polymerase rather than a proofreading function, since it is far from the proofreading 3′–5′ exonuclease domain [41]. However, we have not excluded intraprotein interactions that might affect proofreading.)

We found that unlike for extended deletions between distant repeats, rates of -1 frameshifts in small (2–4 bp) homonucleotide runs were not affected by *pol3-t* [33]. However, there was a considerable potential for *pol3-t*-stimulated genetic change in these runs. The *pol3-t* mutation caused increased rates of frameshifts when MMR was inactivated by the *msh2*, *msh3* mutations disrupting recognition of a mismatch or *pms1* mutation disrupting the second stage of MMR complex formation, suggesting that the frameshift intermediates caused by *pol3-t* were prevented by MMR from becoming mutations. In contrast, MMR was completely unable to prevent large deletions caused by *pol3-t*, although MMR could in part prevent deletions of 7-bp sequences between short repeats also studied in this work (not shown on Fig. 4). We proposed (Fig. 4) that unpaired DNA loops can be formed by replication slippage between repeats and that these loops can be repaired by MMR only if they are small (< 8 nt in our experiments). The absence of an effect of MMR on larger loops (≥ 30 nt in our experiments) could be due to a lack of recognition or because there is no directionality of MMR for such loops. Thus, the ability of *pol3-t* to cause high rates of ARM-associated changes is due to high likelihood of forming mutation intermediate (Premutational structure in Table 1) that is a poor substrate for a second DNA metabolic system, MMR.

Another type of genetic change associated with the short separated repeat category of ARMs is the generation of duplications (Figs. 3 and 5). These duplications are greatly (> 1000 fold) increased in a yeast *rad27*-null mutant and, similar to *pol3-t* stimulated deletions, cannot be prevented by MMR [38].

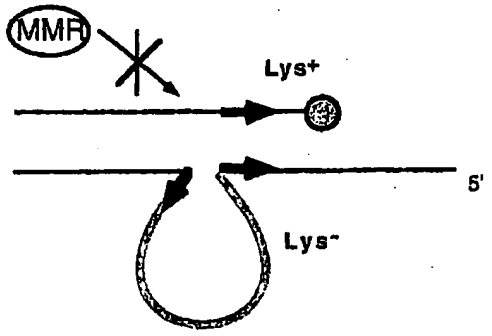


Fig. 4. Looped intermediate arising by replication slippage leads to deletions between short direct repeats. DNA polymerase (gray ball) slippage between two repeats generates an unpaired loop. For the model *LYS2* system in yeast [33,37], slippage between two short (6–7 bp) direct repeats (thick arrows) in the yeast *LYS2* gene separated by either 31 or 61 nt is shown. One of these repeats is at the end of an insert (shown by gray line) that inactivates the *LYS2* gene (Lys^-). Another repeat is in the *LYS2* sequence adjacent to an insert. If the looped intermediate is not repaired back to the initial sequence by MMR, the nascent strand yields a Lys^+ revertant.

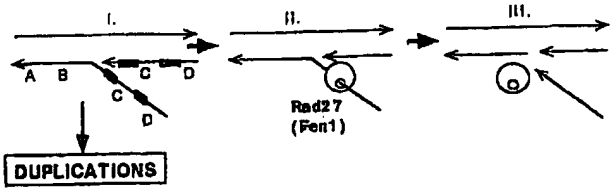


Fig. 5. Defect in removal of a displaced flap can lead to duplications between distant short repeats. Presented is a model describing how a flap is normally removed. (I) Displacement of the 5'-end of the downstream Okazaki fragment generates a 5'-flap containing genetic redundancy (C and D); (II) the Fen1 (Rad27) endonuclease is loaded at the 5'-end of the flap and slides down to the base of the flap; (III) the flap is removed by the Fen1-endonuclease. If the pathway of flap removal is blocked by mutation [38] or by non-canonical DNA structure in the flap (proposed by Gordenin et al. [1]), duplications can occur between direct repeats (gray thick lines) included in the flap by sequence realignment or via end joining if a double-strand break occurs at the base of the flap [38].

The *RAD27* gene (a homologue of human *FEN1*) controls the 5'-flap endonuclease responsible for the removal of displaced flaps at the border between Okazaki fragments in the lagging strand. The unremoved flaps create temporary genetic redundancy and can lead to duplications. In the absence of Rad27 (Fen1), unremoved flaps may be processed via error-prone pathways leading to duplications via flap realignment or via creation of a DSB followed by end-joining (see Fig. 4 in Ref. [38]).

5. Minisatellites

Minisatellite DNAs are tandem repeats where the repeat unit is longer than the repeats in microsatellites. The actual distinction between minisatellites and microsatellites has been arbitrary. Some minisatellites are highly unstable in the human genome [42]. Genetic control of minisatellite instability appeared to be different from that of microsatellites. Sia et al. [26] showed that inactivation of MMR increased up to 6300 fold the rate of changes in tandem short repeats (1–13 bp). However, a MMR defect did not destabilize multiple tandem repeats with longer (16 bp and 20 bp) repeated unit. This suggested that the intermediates leading to deletions and duplications in minisatellite DNAs (e.g., unpaired loops resulting from replication slippage or displaced 5'-flaps) are not substrates for MMR. (It was suggested [22] that microsatellites and minisatellites could be distinguished on the basis of ability of MMR to prevent repeat deletion and duplication. In yeast, such a border would be placed between 13 bp and 16 bp repeat sizes).

As noted above, microsatellite ARMs are highly unstable in MMR deficient human cells, which also makes them a convenient indicator of MMR defects [24,25]. Unlike for microsatellites, no mutator mutations had been reported that could destabilize minisatellites. Based on the observation that *pol3-t* strongly facilitated deletions [33] and *rad27* highly increased duplications [38] associated with random short repeats and that MMR did not prevent these mutations, we proposed that these two mutators could increase deletions and duplications in minisatellite ARMs [43]. In support of this proposal, *pol3-t* and *rad27* were demonstrated to elevate rates of changes in the model yeast minisatellite comprised of three 20 bp repeats (collaborative results of the Petes lab and this lab [44]).

6. Triplet repeats and other small repeats prone to expansion

There is a subset of microsatellite and minisatellite ARMs in the human genome that are capable of not only undergoing small deletions and duplications but also of giving rise to large expansions. Increases in repeat units vary from several to over 100 fold and the probability of expansion increases greatly in long repeats [45–49]. Expansions are primarily observed with the trinucleotide sequences CTG (CAG), CCG (CGG) and GAA (TTC). Recently, expansion was also observed in two minisatellites [50–52]. It is generally acknowledged that repeats that are prone to expansion can form non-canonical DNA structures [48,49,53]. We consider the triplet repeats as ARMs because it is the arrangement of the sequence that is key to expansion.

As discussed above, repeats can become ARMs because they form premutation intermediates that are poor substrates for some DNA metabolic activity. We have proposed that the non-canonical DNA structure that leads to triplet repeat expansion is a poor substrate for enzyme(s) that process the 'flap'

replication intermediates [1]. DNAs can expand if the flap that arises during lagging strand replication is not removed by the Rad27 (Fen1) flap endonuclease (discussed above and in Refs. [1,38,43]). Since the mammalian protein Fen1 can process a flap only if the flap is completely single-stranded [54,55], a secondary structure formed by a flap will be a poor substrate for the this enzyme. We have proposed that the capability of some repeats to form hairpins, triplex DNA or other non-canonical DNA structures in the flap would account for specificity of repeat expansion [1].

Recently, it was demonstrated that *rad27* (*fen1*) deficiency in yeast can cause not only duplications of unique sequences but also small and large expansions in CAG (CTG) repeats [56,57]. We anticipate additional mutators that can increase the likelihood of repeat expansion (Table 1). Mutators could exert their effects by increasing the incidence of structures that are poor substrates in DNA metabolism, such as *pol3-t* which leads to more replication slippage loops, or *rad27* which leads to more flaps. Alternatively, mutators could be in genes whose products are involved in removing or processing replication intermediates even if they assume non-canonical structures. An example of such a mutator is the MMR defect described above that causes greatly increased mutation rates due to lack of removal of the errors that escape DNA-Pol proofreading. Mutators for increased triplet repeat expansion might be due to replication defects that result in a higher incidence of strand displacement in the lagging strand (such as *rad27*) or loss of a proposed exonuclease that acts like Rad27 (Fen1) but its ability to process flaps would not be inhibited by secondary structure.

7. ARMs as sources of human genetic risk

7.1. Genes with ARMs are highly susceptible to change

Although an ARM is only a small sequence within a gene, there are several examples demonstrating that the genetic impact of an ARM can be immense relative to possible changes in the rest of a gene.

Deletions involving distantly separated short repeats or duplications in coding sequences can lead to loss of gene function. Among 100 bp random sequences, approximately 70% are expected to contain a direct repeat of at least 6 bp, suggesting that almost all coding sequences contain this ARM [58]. Thus, they might have a disproportionate effect on genetic change, especially under conditions of altered replication.

Triplet repeat expansion causes inactivation of several genes associated with neurological and muscular function. If the stretch of repeats is long, the probability of expansion in the germ line can reach nearly 100% [59].

Several coding sequences contain repeated peptide motifs encoded by minisatellites with more or less perfect identity between repeats [60–62]. Variations in minisatellite length are common sources of change in encoded protein as described for human and porcine mucin genes [63–66].

Alu repeats are abundant in the human genome and can be associated with rearrangements that lead to disease (see examples in Refs. [67–72]). Based on *Alu*-PCR [73] and indirect physical methods [74,75], many *Alu*-repeats form closely spaced IRs although they are nonidentical. Several examples of closely spaced *Alu*-IRs have been identified during sequencing of the human genome [72,76–79], including some that are nearly palindromic [80]. It is worth noting that very closely spaced *Alu* repeats usually exist in direct orientation. The survey of large number of *Alu*s in the human genome revealed only 47 out of 634 closely spaced (< 21 bp) *Alu* repeats were in the inverted orientation [81]. Such a preference would be expected if very closely spaced long IRs in humans are as unstable as in bacteria and yeast. IRs formed by *Alu* and other genomic repeats could lead to a high frequency of rearrangements especially in mutator backgrounds.

The long homonucleotide ARMs combined with mutators inactivating MMR are especially likely to cause gene inactivation. We directly evaluated the impact of A_{13} and A_{14} homonucleotide runs in yeast. In a mismatch repair deficient (Mmr^-) (*msh2*) strain, mutations that inactivate the *LYS2* gene due to frameshifts in an A_{13} run occur 100 times more frequently than mutations in the rest of the 4179 bp *LYS2* coding sequence [30]. Thus, genes with long homonucleotide runs will form a gene group that is at-risk for inactivation in Mmr^- cells. Since many

tumors are MMR deficient, it is important to know which genes contain long homonucleotide runs in their coding sequences. Some of these genes could be important for cancer progression and for secondary effects of cancer. For example, the coding sequence of the structural gene for human parathyroid hormone-like protein (hPTHrP [82–84]) contains an A_{11} run. Since this protein is associated with hypercalcemia of malignancy, important secondary effect identified with many tumors [85], the controlling gene could be a frequent target for mutation in Mmr^- tumors possibly resulting in the hypercalcination syndrome.

Long homonucleotide runs in coding sequences may represent polymorphisms in human populations. For example, the A_8 run in the APC gene leads to a predisposition to colorectal cancer [86]. The most common allele in the population contains the interrupted run A_3TA_4 and both alleles produce functional protein. In colorectal cancer patients that originally carried the A_8 allele, somatic mutations were often found inside the A_8 run or in the flanking nucleotide. This suggested that the cause of predisposition to colorectal cancer is due to a high likelihood of inactivation of the A_8 allele. Resequencing of interrupted homonucleotide runs in coding sequences of human genes from many individuals can yield information about potentially harmful polymorphisms.

It is interesting that homonucleotide runs as well as various repeats can pose another kind of threat to human health. Pathogenic bacteria can undergo cell surface changes that enables them to escape immuno-surveillance as a result of simple frameshift mutations [87,88]. The cell changes provide for adaptive evolution so that the 'phase changes' which are reversible (by mutation) enable rapid yet relatively stable changes to new variants. Reduction in mismatch repair would increase the likelihood that these ARMs would be mutated.

7.2. Multiple gene inactivation is increased for genes containing ARMs

Phenotypic changes for a particular characteristic may involve mutations in more than one gene so that the likelihood of change is the product of mutation rates for the individual genes. Since genes containing ARMs such as homonucleotide runs are mutation prone, the likelihood of more than one gene being inactivated or altered is much greater for such genes than for genes lacking an ARM. For example, the rate of inactivation of the *LYS2* gene in a Mmr^- (*msh2*) strain is about 10^{-5}; however, the rate is increased to 10^{-3} if the gene contains an A_{13} in frame homonucleotide run [30]. The rate of appearance of double mutants for the group of genes containing the long homonucleotide run would be 10^{-6} as compared to 10^{-10} for most genes which lack such ARMs, suggesting that multiple mutants for these at-risk genes have a reasonable likelihood of occurring.

7.3. Preferential occurrence of premutational DNA structures in ARMs

ARMs, such as homonucleotide runs, could be sites within the genome at which premutational changes occur preferentially. Based on the high mutation rates in long homonucleotide runs in MMR deficient yeast cells, large numbers of mismatches are expected to occur in these ARMs. The human genome contains many long homonucleotide runs in noncoding sequences. For example, there are up to a million *Alu*-repeats in a diploid human cell and most of these contain a long polyA run [74,89]. If human cells have the same high rates of mismatch generation in long homonucleotide runs as yeast (i.e., 10^{-3} for runs containing ≥ 13 As) a thousand mismatches are expected to be generated during every round of replication of the *Alu*s within the human genome. Considerable numbers of mismatches are also expected to be generated in other microsatellite sequences. The number of mismatches in nonreiterated sequences should be much less. Based on rates of forward mutation in genes without long homonucleotide runs in MMR deficient human cells (around 10^{-5} per 1 kb of coding sequence for the HPRT gene [90]) it is expected that around 10 mismatches would be generated during each replication in the $\sim 10^9$ nucleotides of non-microsatellite sequences in the human genome. Thus, mismatches (unpaired loops) in homonucleotide and other microsatellite ARMs are expected to constitute the predominant

substrate for MMR in wild type human cells. Under conditions of limited MMR, they might saturate the MMR system, and in the absence of MMR, they would be primary sites of mutation. If unrepaired mismatches lead to secondary events such as DNA damage signaling response, these ARMs would be a major source of such signals.

7.4. ARMs can amplify the effects of mutators

Mutators that can increase the incidence of certain premutational structures, such as *pol3-t* generated replication slippage loops [33] or unprocessed flaps in the *rad27*-null mutant [38] may have a much greater effect in ARMs sequences. Premutation structures in ARMs are likely to be stabilized or likely to lead to secondary events (e.g., due to presence of small repeats). The *pol3-t* and *rad27* mutators increase rates of repeat-associated deletions [33] and duplications [38], respectively, about 1000 fold, whereas they cause much smaller increases in rates of nonspecific mutations leading to gene inactivation (< 10 fold increase for *pol3-t* [91] and a 50-fold increase for *rad27* [38]). Short quasipalindromes did not show a recombinagenic effect in wild type yeast, but they did increase recombination more than 10-fold in the *pol3-t* background [6]. Mutators that occur due to a defect in the repair of premutational DNA structures can also cause a greater increase of mutation rate in ARMs than in non-ARM DNA. For example, the MMR defect caused a 10,000-fold increase in the mutation rate for long (A_{13} and A_{14}) homonucleotide runs, whereas the mutation rates in other sequences are increased less than 100-fold [30].

7.5. At-risk sequences as ARMs

There are many sites within the genome where specific DNA transactions occur and these might be sights for incorrect changes resulting in genome instability. For example, breakpoints have been identified in lymphoma-associated translocations where the DNA sequence is similar to immunoglobulin switch recombination sites [92–95]. These translocations probably occur as a result of mistakes in the immunoglobulin gene switching recombination. Hotspots for acridine-induced frameshift mutation in T4 phage fit into the category of at-risk consensus sequences. These hotspots, which often correspond to cleavage sites for DNA topoisomerase II, probably result from an aberrant cleavage reaction [96].

8. ARMs control and genome stability

8.1. Using ARMs to find mutators and vice versa

The high potential for ARMs to cause genomic changes can be used as a tool for identifying mutators. Microsatellite ARMs are sensitive indicators of MMR defects in human cell lines and cancer tissues [24,25], as well as microbial systems. The hypermutability of ARMs in mutators can be used to find new mutators, especially if the mutator has only a small overall effect on mutations in the genome. For example, a mutator defect that leaves only 0.1% of mismatches unrepaired in the *lys2-*A_{14} allele described above would still be expected to cause a 10-fold increased mutation rate. This property of long homonucleotide runs has, in fact, been used to find new mutators that can affect MMR. We found that inactivation of the yeast *EXO1* gene (encoding one of redundant 5′ to 3′ exonucleases) caused a 100-fold increase in the mutation rate within an A_{14} run, corresponding to only 1% of mismatches being unrepaired [97], whereas it had only a small effect in other mutation systems. It is unlikely that this gene would have been picked up with other mutation screens, although other approaches have led to identification of *EXO1* interaction with MMR [98]. The long homonucleotide runs have also been useful for studying effects of specific mutations in the *MSH2* gene and the identification of subtle changes in MMR function (unpublished collaborative results of Kunkel and Resnick labs at NIEHS). Based on the high sensitivity of the ARM-based mutator screens we propose that they can be used to identify very weak mutator polymorphisms in the human population.

Just as ARMs can be been used to identify weak mutators, we propose that strong mutators might reveal motifs or consensus sequences that are more prone to spontaneous or damage induced changes. This could be addressed directly by inserting the regions of the human genome into model mutation or recombination substrates and examining them in var-

ious mutator strains. Thus, the mutator strains provide 'sensitized' systems for revealing new ARMs. This approach is likely to prove useful in the study of novel categories of mutators. For example, Glassner et al. [99] and Xiao and Samson [100] showed that overexpressed yeast protein Mag1 (3-methyladenine DNA glycosylase) greatly increases mutation frequencies in *E. coli* and yeast. They proposed that the mutator effect might be due to the increased level of potentially mutagenic basic sites resulting from the release of normal bases by Mag1. This activity might be inefficient in wild type cells expressing natural levels of glycosylase [99]. Such a release was recently demonstrated for *E. coli*, yeast, and human glycosylases [101]. It would be interesting to determine if this activity exhibited sequence or motif specificity.

8.2. Environmental factors that amplify the impact of ARMs

ARMs that affect spontaneous changes in the genome could also be sites of increased instability due to environmental factors. For example, the frameshift mutagen ICR-191 and the carcinogen *N*-2-acetylaminofluorene induces mutation preferentially in homonucleotide or dinucleotide repeats (reviewed in Ref. [102]). Recently, it was demonstrated that minisatellite ARM mutations in cultured mouse cells can be facilitated by the tumor promoter okadaic acid [103]. Alternatively, there might be motifs that only become at-risk when cells are exposed to environmental factors.

Since altered DNA replication (DNA-Pol, flap endonuclease) can increase deletions and recombination, drugs affecting replication may increase ARM-associated genome instability. Agents such as cytosine arabinoside and antifolate drugs, which cause recombination, amplification and other chromosome changes [104–106], would be good candidates. Treatment of cultured human cells with antifolates and several other agents that interfere with DNA replication cause cytologically detectable gaps in the regions of chromosomes called fragile sites. Many of these sites are formed by ARMs prone to expansion, i.e., CCG (CGG) triplet repeats (reviewed in Refs. [45,107] or AT-rich minisatellite [50]. Some fragile sites are associated with chromosome breakage and rearrangements in humans [108,109]. The reactivity of fragile sites to replication inhibitors observed in cytological preparations suggests that fragile sites may contain ARMs prone to induced changes.

9. A general view of ARMs

ARMs potential for causing genomic instability is clearly due to an interplay between ARMs and DNA metabolic systems. With the sequencing of many genomes, including that of humans, and the ability to modify DNA metabolic systems, we anticipate the identification of new ARMs and factors that affect ARMs stability. ARMs may be important components in the response to environmental agents. Investigations of the interaction of ARMs with chemicals that damage DNA or modify the DNA metabolic machinery are likely to reveal novel genetic threats and identify genomic regions particularly prone to environmentally induced changes.

Acknowledgements

We are grateful to R. Slebos, A. Green, K. Lobachev, Tran Hiep for their critical reading of the manuscript and useful discussions.

References

[1] D.A. Gordenin, T.A. Kunkel, M.A. Resnick, Repeat expansion—all in a flap?, Nat. Genet. 16 (1997) 116–118.
[2] R.R. Sinden, DNA Structure and Function, Academic Press, San Diego, 1994, 398 pp.
[3] J. Collins, Instability of palindromic DNA in *Escherichia coli*, Cold Spring Harb. Symp. Quant. Biol. 45 (1981) 409–416, Pt. 1.
[4] J. Collins, G. Volckaert, P. Nevers, Precise and nearly-precise excision of the symmetrical inverted repeats of *Tn5*; common features of recA-independent deletion events in *Escherichia coli*, Gene 19 (1982) 139–146.
[5] D.S. Ehrlich, Illegitimate recombination in bacteria, in: D.E. Berg, M.M. Howe (Eds.), Mobile DNA, American Society for Microbiology, Washington, DC, 1989, pp. 799–832.
[6] D.R. Leach, Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair, Bioessays 16 (1994) 893–900.
[7] E. Akgun, J. Zahn, S. Baumes, G. Brown, F. Liang, P.J. Romanienko, S. Lewis, M. Jasin, Palindrome resolution and recombination in the mammalian germ line, Mol. Cell. Biol. 17 (1997) 5559–5570.

[8] U. DasGupta, K. Weston-Hafer, D.E. Berg, Local DNA sequence control of deletion formation in *Escherichia coli* plasmid pBR322, Genetics 115 (1987) 41–49.

[9] D.E. Berg, C. Egner, B.J. Hirschel, J. Howard, L. Johnsrud, R.A. Jorgensen, T.D. Tlsty, Insertion, excision, and inversion of *Tn5*, Cold Spring Harb. Symp. Quant. Biol. 45 (1981) 115–123, Pt. 1.

[10] C. Egner, D.E. Berg, Excision of transposon *Tn5* is dependent on the inverted repeats but not on the transposase function of *Tn5*, Proc. Natl. Acad. Sci. USA 78 (1981) 459–463.

[11] T.J. Foster, V. Lundblad, S. Hanley-Way, S.M. Halling, N. Kleckner, Three *Tn10*-associated excision events: relationship to transposition and role of direct and inverted repeats, Cell 23 (1981) 215–227.

[12] D.A. Gordenin, K.S. Lobachov, N.P. Degtyareva, A.L. Malkova, E. Perkins, M.A. Resnick, Inverted DNA repeats: a source of eucaryotic genomic instability, Mol. Cell. Biol. 13 (1993) 5315–5322.

[13] D.A. Gordenin, A.L. Malkova, A. Peterzen, V.N. Kulikov, Y.I. Pavlov, E. Perkins, M.A. Resnick, Transposon *Tn5* excision in yeast: the influence of DNA polymerases alpha, delta, epsilon and repair genes, Proc. Natl. Acad. Sci. USA 89 (1992) 3785–3789.

[14] B. Ruskin, G.R. Fink, Mutations in *POL1* increase the mitotic instability of tandem inverted repeats in *Saccharomyces cerevisiae*, Genetics 134 (1993) 43–56.

[15] K.S. Lobachev, B.M. Shor, T.T. Hlep, W. Taylor, J.D. Keen, M.A. Resnick, D.A. Gordenin, Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*, Genetics, 1998, in press.

[16] H. Tran, N. Degtyareva, D. Gordenin, M.A. Resnick, Altered replication and inverted repeats induce mismatch repair-independent recombination between highly diverged DNAs in yeast, Mol. Cell. Biol. 17 (1997) 1027–1036.

[17] M.A. Resnick, The repair of double-strand breaks in DNA: a model involving recombination, J. Theor. Biol. 59 (1976) 97–106.

[18] A. Kuzminov, Collapse and repair of replication forks in *Escherichia coli*, Mol. Microbiol. 16 (1995) 373–384.

[19] H. Bierne, B. Michel, When replication fork stops, Mol. Microbiol. 13 (1994) 17–23.

[20] B. Michel, S.D. Ehrlich, M. Uzest, DNA double-strand breaks caused by replication arrest, EMBO J. 16 (1997) 430–438.

[21] T. Formosa, B. Alberts, DNA synthesis dependent on genetic recombination: characterization of a reaction catalyzed by purified bacteriophage T4 proteins, Cell 47 (1986) 793–806.

[22] E.A. Sia, R.J. Kokoska, M. Dominska, P. Greenwell, T.D. Petes, Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes, Mol. Cell. Biol. 17 (1997) 2851–2858.

[23] J. Dubovsky, V.C. Sheffield, G.M. Duyk, J.L. Weber, Sets of short tandem repeat polymorphisms for efficient linkage screening of the human genome, Hum. Mol. Genet. 4 (1995) 449–452.

[24] M. Perucho, Cancer of the microsatellite mutator phenotype, Biol. Chem. 377 (1996) 675–684.

[25] A. Umar, T.A. Kunkel, DNA-replication fidelity, mismatch repair and genome instability in cancer cells, Eur. J. Biochem. 238 (1996) 297–307.

[26] E.A. Sia, S. Jinks-Robertson, T.D. Petes, Genetic control of microsatellite stability, Mutat. Res. 383 (1997) 61–70.

[27] N. Rampino, H. Yamamoto, Y. Ionov, Y. Li, H. Sawai, J.C. Reed, M. Perucho, Somatic frameshift mutations in the *BAX* gene in colon cancers of the microsatellite mutator phenotype, Science 275 (1997) 967–969.

[28] S. Markowitz, J. Wang, L. Myeroff, R. Parsons, L. Sun, J. Lutterbaugh, R.S. Fan, E. Zborowska, K.W. Kinzler, B. Vogelstein, M. Brattain, J.K.V. Wilson, Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability, Science 2685 (1995) 1336–1338.

[29] R.F. Souza, R. Appel, J. Yin, S. Wang, K.N. Smolinski, J.N. Abraham, T.T. Zou, J.Q. Shi, J. Lei, J. Cottrell, K. Cymes, K. Biden, L. Simms, B. Leggett, P.M. Lynch, M. Fraizer, S.M. Powell, N. Harpaz, H. Sugimura, J. Young, S.J. Meltzer, Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumors, Nat. Genet. 14 (1996) 255–257.

[30] T.H. Tran, J.D. Keen, M. Kricker, M.A. Resnick, D.A. Gordenin, Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants, Mol. Cell. Biol. 17 (1997) 2859–2865.

[31] A. Morrison, A.L. Johnston, L.H. Johnston, A. Sugino, Pathway correcting DNA replication errors in *S. cerevisiae*, EMBO J. 12 (1993) 1467–1473.

[32] A. Morrison, A. Sugino, The 3′–5′ exonucleases of both DNA polymerase δ and ϵ participate in correcting errors of DNA replication in *S. cerevisiae*, Mol. Gen. Genet. 242 (1994) 289–296.

[33] T.-H. Tran, D. Gordenin, M. Resnick, The prevention of repeat-associated deletions in *Saccaromyces cerevisiae* by mismatch repair depends on size and origin of deletions, Genetics 143 (1996) 1579–1587.

[34] B.S. Strauss, D. Sagher, S. Acharya, Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA, Nucleic Acids Res. 25 (1997) 806–813.

[35] L.C. Kroutil, K. Register, K. Bebenek, T.A. Kunkel, Exonucleolytic proofreading during replication of repetitive DNA, Biochemistry 35 (1996) 1046–1053.

[36] R.M. Schaaper, Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*, J. Biol. Chem. 268 (1993) 23762–23765.

[37] M. Meuth, Illegitimate recombination in mammalian cells, in: D.E. Berg, M.M. Howe (Eds.), Mobile DNA, American Society for Microbiology, Washington, DC, 1989, pp. 833–860.

[38] H.T. Tran, N.P. Degtyareva, N.N. Koloteva, A. Sugino, H. Masumoto, D.A. Gordenin, M.A. Resnick, Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the *RAD50* and *RAD52* genes, Mol. Cell. Biol. 15 (1995) 5607–5617.

[39] D.X. Tishkoff, N. Filosi, G.M. Gaida, R.D. Kolodner, A novel mutation avoidance mechanism dependent on *S. cerevisiae rad27* is distinct from DNA mismatch repair, Cell 88 (1997) 253–263.

[40] R.C. von Borstel, R.W. Ord, S.P. Stewart, R.G. Ritzel, G.S. Lee, U.G. Hennig, E.A. Savage, The mutator *mut7-1* of *Saccharomyces cerevisiae*, Mutation Res. 289 (1993) 97–106.

[41] A. Morrison, J.B. Bell, T.A. Kunkel, A. Sugino, Eukaryotic DNA polymerase amino acid sequence required for 3′–5′ exonuclease activity, Proc. Natl. Acad. Sci. USA 88 (1991) 9473–9477.

[42] A.J. Jeffreys, M.J. Allen, J.A. Armour, A. Collick, Y. Dubrova, N. Fretwell, T. Guram, M. Jobling, C.A. May, D.L. Neil, R. Neumann, Mutation processes at human minisatellites, Electrophoresis 16 (1995) 1577–1585.

[43] T.A. Kunkel, M.A. Resnick, D.A. Gordenin, Mutator specificity and disease: looking over the FENce, Cell 88 (1997) 155–158.

[44] R.J. Kokoska, L. Stefanovic, H.T. Tran, M.A. Resnick, D.A. Gordenin, T.D. Petes, Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (*rad27*) and DNA polymerase delta (*pol3-t*), Mol. Cell. Biol. (1998), in press.

[45] C.J. Ashley, S.T. Warren, Trinucleotide repeat expansion and human disease, Annu. Rev. Genet. 29 (1995) 703–728.

[46] S.T. Warren, The expanding world of triplet repeats, Science 271 (1996) 1374–1375.

[47] J.L. Mandel, Breaking the rule of three, Nature 386 (1997) 767–769.

[48] C.T. McMurray, Mechanisms of DNA expansion, Chromosoma 104 (1995) 2–13.

[49] M. Mitas, Trinucleotide repeats associated with human disease, Nucleic Acids Res. 25 (1997) 2245–2254.

[50] S. Yu, M. Mangelsdorf, D. Hewett, L. Hobson, E. Baker, H.J. Eyre, N. Lapsys, P.D. Le, N.A. Doggett, G.R. Sutherland, R.I. Richards, Human chromosomal fragile site *FRA16B* is an amplified AT-rich minisatellite repeat, Cell 88 (1997) 367–374.

[51] D.M. Lalioti, S.S. Hamish, C. Buresi, C. Rosier, A. Bottani, M.A. Morris, A. Malafosse, S.E. Antorakis, Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy, Nature 386 (1997) 847–851.

[52] M.D. Lalioti, H.S. Scott, S.E. Antonarakis, What is expanded in progressive myoclonus epilepsy?, Nat. Genet. 17 (1997) 17.

[53] R.D. Wells, Molecular basis of genetic instability of triplet repeats, J. Biol. Chem. 271 (1996) 2875–2878.

[54] R.A. Bambara, R.S. Murante, L.A. Henricksen, Enzymes and reactions at the eukaryotic DNA replication fork, J. Biol. Chem. 272 (1997) 4647–4650.

[55] M.R. Lieber, The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair, Bioessays 19 (1997) 233–240.

[56] J.K. Schweitzer, D.M. Livingston, Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation, Hum. Mol. Genet. 7 (1998) 69–74.

[57] C.H. Freudenreich, S.M. Kantrow, V.A. Zakian, Expansion and length-dependent fragility of CTG repeats in yeast, Science, 1998, in press.

[58] G.P. Moore, A.R. Moore, L.I. Grossman, The frequency of matching sequences in DNA, J. Theor. Biol. 108 (1984) 111–122.

[59] Y.H. Fu, D.P. Kuhl, A. Pizzuti, M. Pieretti, J.S. Sutcliffe, S. Richards, A.J. Verkerk, J.J. Holden, R.G. Fenwick Jr., S.T. Warren et al., Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox, Cell 67 (1991) 1047–1058.

[60] J. Heringa, P. Argos, A method to recognize distant repeats in protein sequences, Proteins 17 (1993) 341–391.

[61] J. Heringa, The evolution and recognition of protein sequence repeats, Comput. Chem. 18 (1994) 233–243.

[62] N. Matsushima, C.E. Creutz, R.H. Kretsinger, Polyproline, beta-turn helices. Novel secondary structures proposed for the tandem repeats within rhodopsin, synaptophysin, synexin, gliadin, RNA polymerase II, hordein, and gluten, Proteins 7 (1990) 125–155.

[63] J.L. Desseyn, V. Guyonnet-Duperat, N. Porchet, J.P. Aubert, A. Laine, Human mucin gene *MUC5B*, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p15.5 gene family, J. Biol. Chem. 272 (1997) 3168–3178.

[64] W.S. Pratt, I. Islam, D.M. Swallow, Two additional polymorphisms within the hypervariable *MUC1* gene: association of alleles either side of the VNTR region, Ann. Hum. Genet. 60 (1996) 21–28.

[65] V. Shankar, M.S. Gilmore, R.C. Elkins, G.P. Sachdev, A novel human airway mucin cDNA encodes a protein with unique tandem-repeat organization, Biochem. J. 300 (1994) 295–298.

[66] A.E. Eckhardt, C.S. Timpte, A.W. DeLuca, R.L. Hill, The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin, J. Biol. Chem. 272 (1997) 33204–33210.

[67] B. Calabretta, D.L. Robberson, H.A. Barrera-Saldana, T.P. Lambrou, G.F. Saunders, Genome instability in a region of human DNA enriched in *Alu* repeat sequences, Nature 296 (1982) 219–225.

[68] A. Jalanko, T. Manninen, L. Peltonen, Deletion of the C-terminal end of aspartylglucosaminidase resulting in a lysosomal accumulation disease: evidence for a unique genomic rearrangement, Hum. Mol. Genet. 4 (1995) 435–441.

[69] M.A. Lehrman, W.J. Schneider, T.C. Sudhof, M.S. Brown, J.L. Goldstein, D.W. Russell, Mutation in LDL receptor: *Alu–Alu* recombination deletes exon encoding transmembrane and cytoplasmic domains, Science 227 (1985) 140–146.

[70] M.A. Lehrman, D.W. Russell, J.L. Goldstein, M.S. Brown, *Exon–Alu* recombination deleted 5 kilobases from the low density lipoprotein receptor, producing a null phenotype in familial hypercholesterolemia, Proc. Natl. Acad. Sci. USA 83 (1986) 3679–3683.

[71] M.A. Lehrman, D.W. Russell, J.L. Goldstein, M.S. Brown, *Alu–Alu* recombination deletes splice acceptor from the low density lipoprotein receptor in a subject with familial hypercholesterolemia, J. Biol. Chem. 262 (1987) 3354–3361.

[72] R.A. Macina, F.G. Barr, N. Galili, H.C. Riethman, Genomic organization of the human *PAX3* gene: DNA sequence analysis of the region disrupted in alveolar rhabdomyosarcoma, Genomics 26 (1995) 1–8.

[73] D.L. Nelson, S.A. Ledbetter, L. Corbo, M.F. Victoria, R. Ramirez-Solis, T.D. Webster, D.H. Ledbetter, C.T. Caskey, *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources, Proc. Natl. Acad. Sci. USA 86 (1989) 6686–6690.

[74] P.L. Deininger, SINES: short interspersed repeated DNA elements in higher eukaryote, in: D.E. Berg, M.M. Howe (Eds.), Mobile DNA, American Society for Microbiology, Washington, DC, 1989, pp. 619–636.

[75] P.L. Deininger, C.W. Schmid, An electron microscope study of the DNA sequence organization of the human genome, J. Mol. Biol. 106 (1976) 773–790.

[76] H. Kiyosawa, P.F. Chance, Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot, Hum. Mol. Genet. 5 (1996) 745–753.

[77] Y. Koda, M. Soejima, B. Wang, H. Kimura, Structure and expression of the gene encoding secretor-type galactoside 2-alpha-L-fucosyltransferase (FUT2), Eur. J. Biochem. 246 (1997) 750–755.

[78] M. Pauly, I. Kayser, M. Schmitz, F. Ries, F. Hentges, M. Dicato, The human *mdr1* (multidrug-resistance) gene harbours a long homopyrimidine–homopurine sequence next to a cluster of *Alu* repeated sequences in intron 14, Gene 153 (1995) 299–300.

[79] I. Unfried, B. Entler, R. Prohaska, The organization of the gene (*EPB72*) encoding the human erythrocyte band 7 integral membrane protein (protein 7.2b), Genomics 30 (1995) 521–528.

[80] J.H. Hanke, J.E. Hambor, P. Kavathas, Repetitive *Alu* elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer, J. Mol. Biol. 246 (1995) 63–73.

[81] J. Jurka, Origin and evolution of *Alu* repetitive elements, in: R.J. Maraia (Ed.), The Impact of Short Interspersed Elements (SINEs) on the Host Genome, Springer, New York, 1995, pp. 25–42.

[82] R.V. Campos, L. Zhang, D.J. Drucker, Differential expression of RNA transcripts encoding unique carboxy-terminal sequences of human parathyroid hormone-related peptide, Mol. Endocrinol. 8 (1994) 1656–1666.

[83] J. Southby, L.M. Murphy, T.J. Martin, M.T. Gillespie, Cell-specific and regulator-induced promoter usage and messenger ribonucleic acid splicing for parathyroid hormone-related protein, Endocrinology 137 (1996) 1349–1357.

[84] T. Yasuda, D. Banville, G.N. Hendy, D. Goltzman, Characterization of the human parathyroid hormone-like peptide gene. Functional and evolutionary aspects, J. Biol. Chem. 264 (1989) 7720–7725.

[85] T.J. Martin, J.M. Moseley, M.T. Gillespie, Parathyroid hormone-related protein: biochemistry and molecular biology, Crit. Rev. Biochem. Mol. Biol. 26 (1991) 377–395.

[86] S.J. Laken, G.M. Petersen, S.B. Gruber, C. Oddoux, H. Ostrer, F.M. Giardiello, S.R. Hamilton, H. Hampel, A. Markowitz, D. Klimstra, S. Jhanwar, S. Winawer, K. Offit, M.C. Luce, K.W. Kinzler, B. Vogelstein, Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC, Nat. Genet. 17 (1997) 79–83.

[87] E.R. Moxon, P.B. Rainey, M.A. Nowak, R.E. Lenski, Adaptive evolution of highly mutable loci in pathogenic bacteria, Curr. Biol. 4 (1994) 24–33.

[88] K.W. Deitsch, E.R. Moxon, T.E. Wellems, Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections, Microbiol. Mol. Biol. Rev. 61 (1997) 281–293.

[89] E.P. Economou, A.W. Bergen, A.C. Warren, S.E. Antonarakis, The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome, Proc. Natl. Acad. Sci. USA 87 (1990) 2951–2954.

[90] W.E. Glaab, K.R. Tindall, Mutation rate at the *hprt* locus in human cancer cell lines with specific mismatch repair-gene defects, Carcinogenesis 18 (1997) 1–8.

[91] D.A. Gordenin, Y.Y. Proscyavichus, A.L. Malkova, M.V. Trofimova, A. Peterzen, Yeast mutants with increased transposon *Tn5* excision, Yeast 7 (1991) 37–50.

[92] C.M. Croce, Role of chromosome translocations in human neoplasia, Cell 49 (1987) 155–156.

[93] L.R. Finger, R.C. Harvey, R.C. Moore, L.C. Showe, C.M. Croce, A common mechanism of chromosomal translocation in T- and B-cell neoplasia, Science 234 (1986) 982–985.

[94] F.G. Haluska, Y. Tsujimoto, G. Russo, M. Isobe, C.M. Croce, Molecular genetics of lymphoid tumorigenesis, Prog. Nucleic Acid Res. Mol. Biol. 36 (1989) 269–280.

[95] S.P. Piccoli, P.G. Calmi, M.D. Cole, A conserved sequence at c-*myc* oncogene chromosomal translocation breakpoints in plasmacytomas, Nature 310 (1984) 327–330.

[96] L.S. Ripley, J.S. Dubins, J.G. deBoer, D.M. DeMarini, A.M. Bogerd, K.N. Kreuzer, Hotspot sites for acridine-induced frameshift mutations in bacteriophage T4 correspond to sites of action of the T4 type II topoisomerase, J. Mol. Biol. 200 (1988) 665–680.

[97] T.H. Tran, M.A. Resnick, D.A. Gordenin, The $5' \rightarrow 3'$ exonucleases *Exo1* and *Rad27* along with the $3' \rightarrow 5'$ exonuclease of DNA polymerase e participate in mismatch repair in *Saccharomyces cerevisiae*, 1998, submitted.

[98] D.X. Tishkoff, A.L. Boerger, P. Bertrand, N. Filosi, G.M. Gaida, M.F. Kane, R.D. Kolodner, Identification and characterization of *Saccharomyces cerevisiae EXO1*, a gene encoding an exonuclease that interacts with *MSH2*, Proc. Natl. Acad. Sci. USA 94 (1997) 7487–7492.

[99] B.J. Glassner, L.M. Posnick, L.D. Samson, The influence of DNA glycosylases on spontaneous mutation, Mutation Res. 400 (1998) 33–44.

[100] W. Xiao, L. Samson, In vivo evidence for endogenous

DNA alkylation damage as a source of spontaneous mutation in eukaryotic cells, Proc. Natl. Acad. Sci. USA 90 (1993) 2117–2121.

[101] K.G. Berdal, R.F. Johansen, E. Seeberg, Release of normal bases from intact DNA by a native DNA repair enzyme, EMBO J. 17 (1998) 363–367.

[102] L.S. Ripley, Frameshift mutation: determinants of specificity, Annu. Rev. Genet. 24 (1990) 189–213.

[103] H. Nakagama, S. Kaneko, H. Shima, H. Inamori, H. Fukuda, R. Kominami, T. Sugimura, M. Nagao, Induction of minisatellite mutation in NIH 3T3 cells by treatment with the tumor promoter okadaic acid, Proc. Natl. Acad. Sci. USA 94 (1997) 10813–10816.

[104] S. Grant, Ara-C: cellular and molecular pharmacology, Adv. Cancer Res. 72 (1998) 197–233.

[105] B.J. Barclay, B.A. Kunz, J.G. Little, R.H. Haynes, Genetic and biochemical consequences of thymidylate stress, Can. J. Biochem. 60 (1982) 172–184.

[106] B. Kamen, Folate and antifolate pharmacology, Semin. Oncol. 24 (1997) S1830–S1839.

[107] G.R. Sutherland, R.I. Richards, The molecular basis of fragile sites in human chromosomes, Curr. Opin. Genet. Dev. 5 (1995) 323–327.

[108] C. Jones, L. Penny, T. Mattina, S. Yu, E. Baker, L. Voullaire, W.Y. Langdon, G.R. Sutherland, R.I. Richards, A. Tunnacliffe, Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2, Nature 376 (1995) 145–149.

[109] H. Inoue, H. Ishii, H. Alder, E. Snyder, T. Druck, K. Huebner, C.M. Croce, Sequence of the FRA3B common fragile region: implications for the mechanism of FHIT deletion, Proc. Natl. Acad. Sci. USA 94 (1997) 14584–14589.

Pergamon

PII: S0965-1748(96)00022-7

Insect Biochem. Molec. Biol. Vol. 26, Nos 8–9, pp. 739–753, 1996

0965-1748/96 $15.00 + 0.00

Mini-Review

Transposable Elements and Gene Transformation in Non-Drosophilid Insects

DAVID A. O'BROCHTA,*‡ PETER W. ATKINSON†

Received 1 December 1995; revised and accepted 25 March 1996

This review summarizes recent data on the development of non-drosophilid insect transformation systems. The discussion focuses on one particular approach to developing transformation systems that relies on the use of short inverted repeat-type transposable elements analogous to that employed for *Drosophila melanogaster* transformation. Representatives from four families of short inverted repeat-type transposable elements have been shown to either act as non-drosophilid gene vectors or to have the ability to transpose accurately when introduced into non-host insect cells. *Minos*, a member of the *Tcl* family of elements isolated originally from *D. hydei* has been successfully used as a germline transformation vector in the Medfly, *Ceratitis capitata*. *Hermes*, a member of the *hAT* family of elements isolated originally from *Musca domestica* has been successfully used as a gene transformation vector in *D. melanogaster* and has a host range that appears to include culicids. *hobo*, another member of the *hAT* family of elements isolated from *D. melanogaster* also has a broad host range that includes tephritid fruitflies. *mariner(Mos)*, a member of the *mariner* family of elements isolated from *D. mauritiana* can transpose in calliphorids. Finally, *piggyBac/IFP2*, a member of the TTAA-specific family of elements isolated from *Trichoplusia ni* can transpose when introduced into *Spodoptera frugiperda* cells. Although routine transformation of insects other than *D. melanogaster* is not possible it is clear that the raw materials for the development of efficient transformation systems are now available. Copyright © 1996 Elsevier Science Ltd

Transposable elements Gene vectors Transformation Transgenic insects *P*-Elements *mariner hobo Hermes Minos Tc1 Musca domestica Drosophila melanogaster Aedes aegypti Lucilia cuprina—Cochliomyia hominivorax Ceratitis capitata Bactrocera tryoni* Sterile insect technique Biological control

INTRODUCTION

The technology for creating transgenic non-drosophild insects will have many applications. Some applications, such as manipulating *Apis mellifera* or *Bombyx mori* breeding stocks, are perfectly analogous to the uses of transgenic technology in plant and livestock breeding programs. Other proposed applications, such as replacing populations of pest insects with a genetically engineered non-pest strain, are novel and controversial (Collins, 1994; Curtis, 1994; Spielman, 1994). Recent advances in research to develop this technology has resulted in the development of the basic components required for efficient and routine introduction of genes into the genomes of insects of economic and medical importance. To a very limited extent these components have been assembled and used successfully.

There are numerous strategies for transforming metazoans but only a few are likely to find widespread application in insect science. Random integration of DNA introduced into germ- or stem-cells provides a workable strategy for creating transgenic mammals (Maclean, 1995) but will only be of limited value in producing transgenic arthropods until efficient selection methods are developed for recognizing rare transgenic insects. Similarly, retroviruses play a major role in efforts to create transgenic animals and animal cells with the exception of insects (Maclean, 1995). Until recently the use of

*Center for Agricultural Biotechnology, University of Maryland, Biotechnology Institute, College Park, MD 20742, U.S.A.
†CSIRO Division of Entomology, Clunies Ross Street, Black Mountain, Canberra, ACT 2601, Australia
‡Author for correspondence. Tel.: 301-405-7680, Fax: 301-314-9075, email: dol4@umail.umd.edu

retroviral vectors was not an option for insect scientists because of the absence of retroviruses that originated from or functioned in insects. The recent discovery of the first invertebrate (*gypsy*) (Kim *et al.*, 1994) and the development of pantropic pseudotyped mammalian retroviruses, that have expanded host ranges (Yee *et al.*, 1994), now permit the construction of retroviral insect vectors to be seriously explored. A completely novel strategy for genetically manipulating the phenotype of insects involves genetically manipulating microbial symbionts that inhabit various insect tissues (Beard *et al.*, 1993). Such insects might be referred to as paratransgenic insects. One strategy, however, that is certain to be generally useful in producing transgenic insects, and is the exclusive focus of this review (Table 1), is the construction of insect gene vectors from transposable elements containing short inverted terminal repeats.

Transposable elements include a diverse collection of genetic elements which have in common the ability to promote recombination reactions that result in the movement of the element from one location in the genome to another. Structurally and mechanistically this is accomplished in a variety of ways. An operational classification system that has proven convenient in discussing transposable elements divides them into two classes depending upon the general features of the mechanisms of movement (Finnegan, 1989). Class I elements are those that transpose by reverse transcription of an RNA intermediate, similar to retroviral integration. Class II elements appear to transpose directly from DNA to DNA. Class II elements are usually of small to moderate length (less than 10 kb). A subclass of these elements have in common the presence of terminal sequences that form inverted repeats. These terminal inverted repeats are usually (but not always) less than 50 base pairs and can be as few as eight base pairs. Members of this subclass of elements are called short inverted repeat-type elements and have proven most useful in constructing gene vectors for *Drosophila* and are becoming increasingly important for plant genome manipulation (Hehl, 1994; Walbot, 1992).

Short inverted repeat-type elements have two major functional components, a transcription unit that encodes a protein (transposase) required for transpositional recombination and the terminal inverted repeats of the element including 100 bases or more of terminus-flanking DNA. Although not well characterized in eukaryotes, transposases are probably recombinases. Likewise the exact role of the terminal repeats of this class of element is not known but they appear to serve as pointers, directing the recombinase and/or other factors to the proper position on the element to initiate recombination. Short inverted-repeat type elements are particularly amenable to being converted to vectors because the two components of the element can function *in trans*. That is, transposase produced by one element can act on the terminal sequences of another element of the same type and promote its transposition. Consequently, attaching the terminal sequences to any other non-transposable element sequence results in the creation of a chimeric transposable element capable of transposing when the appropriate transposase is present. Trans-mobilization is a useful feature of this type of transposable element because it permits the investigator to stabilize the movement of the element following integration by removing the source of transposase. These characteristics were first exploited in eukaryotes for the purposes of gene vector development using the *P*-element from *Drosophila melanogaster* (Rubin and Spradling, 1982; Spradling and Rubin, 1982) and we refer to this as the *P*-element paradigm.

The *P*-element paradigm for genetic transformation of *D. melanogaster* can be simply described as one that relies on a modified short inverted repeat-type transposable element in which the transposase function is provided *in trans* to the essential terminal sequences. The terminal sequences are attached to the gene or DNA sequence to be integrated, forming a chimeric transposable element (now called a vector). Upon entry into a nucleus, transposase promotes the cutting and joining of the vector to a chromosome of the host, resulting in chromosomal integration of the vector. Because movement of the vector does not involve an RNA intermediate, the types of sequences that can be included in the vector are not severely limited and may include introns. A particularly important feature of this paradigm is the

TABLE 1. Short inverted repeat type-transposable elements capable of transposition in non-host species

Element	Host Species	Non Host Species	References
hobo	*D. melanogaster*	*M. domestica*[1] *B. tryoni*[1] *H. armigera*[2] *D. virilis*[3]	[1]O'Brochta *et al.*, 1994; [2]Pinkerton *et al.*, 1996; [3]Lozovskaya *et al.*, 1996
Hermes	*M. domestica*	*D. melanogaster*[1] *B. tryoni*[5] *C. macellaria*[5] *L. cuprina*[5] *Ae. aegypti*[5]	[1]O'Brochta *et al.*, 1996; [5]Sarkar et al. unpublished
Minos	*D. hydei*	*D. melanogaster*[6] *C. capitata*[7]	[6]Loukeris *et al.*, 1995a; [7]Loukeris *et al.*, 1995b
mariner	*D. mauritiana*	*D. melanogaster*[8] *L. cuprina*[9]	[8]Lidholm *et al.*, 1993; [9]Coates et al., unpublished
P	*D. melanogaster*	*D. hawaiiensis*[10] *D. simulans*[11]	[10]Brennan *et al.*, 1984; [11]Scavarda and Hartl, 1984
piggyBac/IFP2	*T. ni*	*S. frugiperda*[12]	[12]Fraser *et al.*, 1995

use of host strains of *D. melanogaster* that do not contain endogenous sources of the appropriate transposase or forms of the element that might inhibit vector integration. This paradigm for gene transfer insures that the integrated vector will not continue to transpose and will not be lost through excision. The existence of *D. melanogaster* strains lacking specific families of transposable elements has been critical in the successful development of gene-transfer technology in this species and has permitted the development of additional vector-based technologies that would not have been possible if stability could not be assured.

The *P*-element has proven useful in manipulating the genome of *D. melanogaster* in ways that go beyond its straightforward use as a gene vector and a simple tool for integrating and expressing foreign DNA. Two of the more notable applications are transposon tagging (Bingham *et al.*, 1981; Searles *et al.*, 1982) and enhancer trapping (Bellen *et al.*, 1989). These are powerful methods using modified gene vectors for identifying and isolating genes. The diversity and effectiveness of tools that have been developed using *P*-elements to investigate the biology of *D. melanogaster* illustrates the utility of short inverted repeat-type transposable elements and justifies current efforts to follow the *P*-element paradigm so that analogous tools can be developed for other insects. Of course, the effectiveness of these tools in *D. melanogaster* is due, in part, to our sophisticated understanding of *Drosophila* genetics and our ability to perform elegant genetic experiments in this species. Consequently, the development and use of gene transfer and related technologies will also require an increase in our efforts to develop a foundation of genetics for non-drosophilid insects. In those insects where there already exists a foundation of genetic information, gene transfer technology is likely to have a rapid, positive impact that will begin in the laboratory and may extend to the field.

PROGRESS TOWARDS FUNCTIONAL NON-DROSOPHILID INSECT VECTOR SYSTEMS

The power and versatility of the *P*-element paradigm has fueled efforts to develop analogous systems for non-drosophilid insects. These efforts have proceeded in two general directions. The first has relied on exploiting transposable elements that were isolated from *D. melanogaster* or related species and testing their abilities to excise, transpose and act as gene vectors in various non-drosophilid insects. The second strategy has involved identification and isolation of new transposable elements from non-drosophilid insects that might be converted to gene vectors. Both strategies have proven successful and operative vectors are now available for a small number of species.

THE P-ELEMENT FAMILY

Description

P elements are the prototypical short inverted repeat-type element in eukaryotes; they are almost 3 kb in length, encode a single protein called transposase and have terminal inverted repeat sequences of 31 nucleotides. The discovery, isolation and characterization of this element has been reviewed elsewhere (Engels, 1989). While *P*-elements provide us with an example of how an effective insect transformation system might be constructed and applied they are currently of only minor interest to those interested in non-drosophilid insect gene vectors. The *P*-element's diminishing significance to current efforts to develop non-drosophilid transformation technology is due to the limited mobility properties of these elements in non-drosophilid species.

Mobility in the host

The *P*-element's mobility in *D. melanogaster* is well studied and documented (Engels, 1989). Movement of *P*-elements is strain-dependent with certain strains repressing almost all *P* movement and others supporting high rates of *P*-element excision and transposition. Within *D. melanogaster*, *P*-elements are effective agents for genome manipulation because of their rates of movement and our abilities to regulate them. Details of these mobility properties can be found elsewhere (Rio, 1991).

Mobility in other insects

The *D. melanogaster P*-element has been successfully used as a transformation vector in the closely related species, *D. simulans* (Scavarda and Hartl, 1984) and a more distantly related drosophilid *D. hawaiiensis* (Brennan *et al.*, 1984). Neither of these species possess endogenous *P*-elements and, in both cases, *P*-elements could undergo transpositions in subsequent generations suggesting that both the transposase and host-supplied factors were present and capable of functioning in these species.

All efforts to employ *P*-elements as gene vectors in non-drosophilid insect species have failed. The successful creation of transgenic non-drosophilid insects using *P*-element vector-containing plasmids has been reported but in no case has the integrated DNA consisted of only sequences precisely delimited by the termini of *P*-elements (McCrane *et al.*, 1988; Miller *et al.*, 1987; Morris *et al.*, 1989). Most efforts to use *P*-elements in non-drosophilids have remained unpublished because transgenic insects were not successfully produced and no evidence of movement was obtained. References to these efforts can be found occasionally [see Walker (1989) for mention of efforts to transform *Locusta migratoria* and *Apis mellifera*] indicating that between 1985 and 1990 there was considerable effort to use *P*-elements in non-drosophilids. The silence that followed those efforts is evidence of their failure. In addition, using *in vivo* plasmid-based recombination assays the excision or transpo-

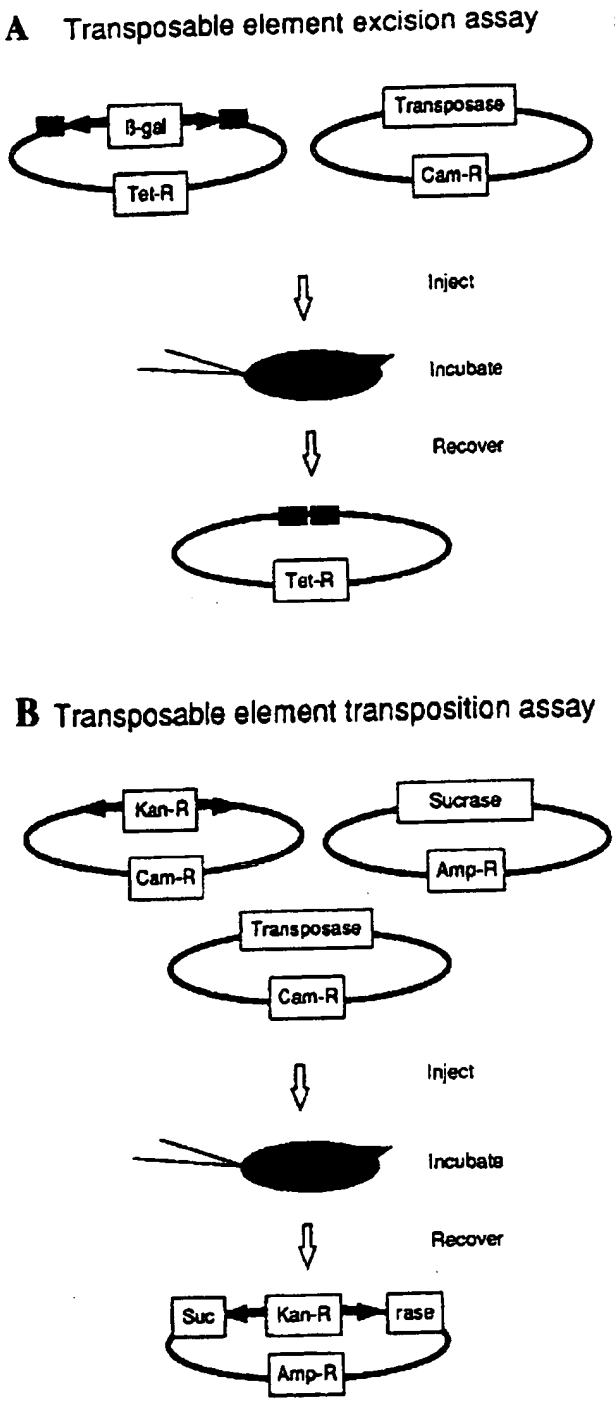


FIGURE 1. Transient transposable element mobility assays performed *in vivo*. (A) Transposable element excision assays. The assay shown is designed to monitor the excision of a *hobo* element containing the gene encoding β-galactosidase although other genetic markers could be used. Excision indicator and helper plasmids are coinjected into preblastoderm insect embryos which are then allowed to develop. Plasmids are recovered and recombinant plasmids resulting from element excision are identified by their unique genotype. By omitting the plasmid encoding transposase the assay can be used to look for the presence of endogenous transposase-like activity in alternative insect hosts. For full details see Atkinson *et al.*, 1993, *Amp-R* = gene encoding ampicillin resistance; *Tet-R* = gene encoding tetracycline resistance. (B) Transposable element transposition assay. The assay shown is designed to monitor the transposition of a *hobo* element containing a gene encoding resistance to kanamycin (*Kan-R*) from the donor plasmid to a target plasmid. Transposition is mediated by the transposase encoded by the helper plasmid. In this version of the assay, the target plasmid contains the *sucraseRB* gene of *Bacillus subtilis*, inactivation of which can be selected for when *Escherichia coli* cells containing it are plated on appropriate media. Transposition is confirmed by determining the DNA sequence of the two junction points of *hobo* transposable element integration into the *sacRB* gene. For full details see O'Brochta *et al.*, 1994. This strategy for assessing transposable element mobility can be employed for any short inverted repeat-type element and as described in the text, related assays are now available for *Hermes* and *mariner* elements.

sition of *P*-elements (Fig. 1) following their introduction into non-drosophilid host insects was never detected (O'Brochta and Handler, 1988). Analysis of the *P*-element's ability to excise in non-hosts, including non-drosophilids, using plasmid-based excision assays revealed a decreased ability to support movement as phylogenetic distance from *D. melanogaster* increased. All non-drosophilid insects tested were unable to support *P*-element excision (O'Brochta and Handler, 1988). Collectively, these data, suggested that, in the absence of significant modification, *P*-elements were unlikely to be useful as transformation vectors in non-drosophilid insect species.

Related elements in other insects

P-elements have been found in a number of drosophilid species including *Scaptomyza pallida*, *D. paulistorum*, *D. bifasciata*, *D. nebulosa* and *D. willistoni* (Anoxolebehere and Periquet, 1987; Daniels *et al.*, 1984, 1990; Hagemann *et al.*, 1990; Lansman *et al.*, 1987). Their discontinuous distribution within this family, together with the high degree of sequence similarity observed among elements from different species suggests that horizontal transfer has, in part, been responsible for the distribution of *P*-elements in the Drosophilidae (Clark *et al.*, 1994; Kidwell, 1992). Until recently the distribution of *P*-elements and *P*-element-related sequences was believed to be rather limited. However, *P*-element-related sequences have been found in *Lucilia cuprina* (Perkins and Howells, 1992) and other cycloraphous diptera (H. Robertson, personal communication).

THE *hobo* OR *hat* ELEMENT FAMILY

Description

The *hobo* element was identified in *D. melanogaster* as a factor causing genetic instabilities when certain strains were hybridized (Lim *et al.*, 1983; Stamatis *et al.*, 1981). Similar instabilities had been found associated with the movement of *P*-elements. The physical isolation of *hobo* occurred during the investigation of glue-protein genes of *D. melanogaster* and was discovered as an insertion element in a spontaneous allele of Sgs-4 (McGinnis *et al.*, 1983).

hobo is a typical short inverted repeat-type element with a copy number of 0–50 per genome, depending on

the strain. It is approximately 3 kb in length with 12 bp terminal inverted repeats, creates 8 bp duplications at the points of integration and probably encodes a single protein (transposase) (Blackman and Gelbart, 1989). Analysis of an autonomous *hobo* element (HFL1) revealed sequence similarities with *Ac* from *Zea mays* (corn) and *Tam3* from *Antirrhinum majus* (snapdragon) (Calvi *et al.*, 1991; Feldmar and Kunze, 1991). The DNA sequence of HFL1 differs from the original description of the *hobo* element ($hobo_{108}$) in five regions, most significantly in that the transposase open reading frame is extended by 42 amino acids at the carboxy end of the polypeptide. The similarities among *hobo*, *Ac* and *Tam3* are largely confined to three regions of the transposase open reading frame and consist of blocks of approximately 50 amino acids that are 35% identical and 50% similar to corresponding regions of *Ac* and *Tam3*. These data suggest that *hobo*, *Ac* and *Tam3* are related members of a family of transposable elements and for convenience we refer to this as the *hAT* family.

Although within the family Drosophilidae *hobo* has a rather limited distribution, the *hAT* family of transposable elements appears widespread in insects (see next section). Of 134 species of *Drosophila* analyzed, *hobo* sequences were found in only the *melanogaster* species group and confined to the *melanogaster* and *montium* subgroups (Daniels *et al.*, 1990). As with *P*-elements there were strains of *melanogaster* that appeared to completely lack the *hobo* element and these were usually older laboratory strains. These data and a comparison of the *hobo* elements found in other *Drosophila* species has led to the hypothesis that *hobo* has been recently (within the last 100 years) introduced into *D. melanogaster* by horizontal transfer (Simmons, 1992).

Mobility in the host

Parallels between structure, distribution and genetics of *P*- and *hobo* elements suggested that *hobo* too could be converted to a useful gene vector. Following vector design, use and detection strategies that precisely paralleled the *P*-element paradigm, Blackman *et al.* (1989) used *hobo* as a gene vector in *D. melanogaster*. *hobo*'s performance was comparable to that of *P*-elements' with approximately 20% or more of the fertile adults developing from embryos injected with the vector producing transgenic progeny. Transposition involved only sequences delimited by the terminal inverted repeats of the element and an 8 bp direct duplication of the integration site was created. Furthermore, *hobo* mobility is confined predominantly to the germline by virtue of tissue specific expression of the *hobo* transposase gene (Calvi and Gelbart, 1994). While *hobo* integration rates are comparable to that observed using *P*-elements, insertion site preferences of these two elements are distinct. The spatial distribution of *hobo* and *P*-element integration sites within the *Drosophila* genome overlapped, but there were chromosomal regions that were preferred by only one of the elements (Smith *et al.*, 1993). This difference in insertion site preference is a useful characteristic of *hobo* vectors because more comprehensive transposon mutagenesis efforts can now be undertaken in *D. melanogaster*.

Although little experimental work on the mechanism of *hobo* movement exists, recent analysis of *hobo* movement in *D. melanogaster* revealed excision reaction-products similar to those recovered following *Ac* and *Tam3* excision in their respective hosts (Atkinson *et al.*, 1993). This similarity supports the hypothesis that *Ac*, *Tam3* and *hobo* are related members of a family of elements that move using a common mechanism.

Mobility in non-hosts

The resemblance of *hobo* to *Ac* and *Tam3* suggested that *hobo* may share additional properties with *Ac* and *Tam3*, such as the ability to transpose when introduced into species which do not normally contain the element (non-host species). The *Ac* transposable element system from maize has an almost unrestricted ability to function in non-host plant cells. *Ac*, although originally isolated from maize, can excise and/or transpose when introduced into many species of monocot- and dicotyledonous plants including tobacco, tomato and petunia (Baker *et al.*, 1986; Chuck *et al.*, 1993; Peterson and Yoder, 1993). *Ac*'s ability to function in non-host species has led to its use as an effective transposon tagging agent in *Arabidopsis thaliana* and other plant species (Aarts *et al.*, 1993; Hehl, 1994). The first direct demonstration of *hobo*'s ability to function in a non-host insect was reported by Atkinson *et al.* (1993). They demonstrated that *hobo* could *excise* from plasmids introduced into and transiently maintained extrachromosomally in cells of developing *Musca domestica* (housefly) embryos (Fig. 1). Similar results were reported in distantly related *Drosophila* species (Handler and Gomez, 1995) and currently the range of species in which *hobo* has been shown to excise includes the mosquitos, *Aedes aegypti* and *Aedes australis* (A. Sarkar, K. Yardley, K. Saville, D. O'Brochta, A. James and P. Atkinson, unpublished results), and the lepidopteran, *Helicoverpa armigera* (Pinkerton *et al.*, 1996).

hobo excision is a good indicator of the presence of *hobo* transposase or transposase-like activity but experiments that directly measure *hobo transposition* are more valuable in determining the potential of the element to serve as a vector. O'Brochta *et al.* (1994) used a modification of the excision assay to directly measure *hobo* transposition in embryos of host and non-host species (Fig. 1). In *D. melanogaster* *hobo* transposition was transposase-dependent, involved only DNA delimited by the 12 bp inverted terminal repeats of the *hobo* element and resulted in the creation of an 8 bp duplication of the insertion site. The consensus integration site was similar to genomic integration sites reported by Streck *et al.* (1986) in *D. melanogaster*. Therefore the plasmid-based *hobo* transposition assay is a reliable indicator of element

behavior and can be used to assess *hobo* mobility in insects other than *Drosophila*.

Interplasmid transposition assays in a number of diverse non-drosophilid insect species including the housefly, *M. domestica* (family, Muscidae), the Queensland fruitfly, *Bactrocera tryoni* (family, Tephritidae) and the corn earworm, *H. armigera* (order, Lepidoptera) demonstrated *hobo's* ability to *transpose* accurately in non-host insects (O'Brochta *et al.*, 1994; Pinkerton *et al.*, 1996). The frequency of *hobo* transposition in non-drosophilids was less than that observed in *D. melanogaster* under similar conditions but, as in *D. melanogaster*, transposition involved only sequences precisely delimited by the terminal inverted repeats and integration occurred into 8 bp target sites resembling those used in the host species. These data demonstrate the ability of *hobo* to function in widely diverged non-host species and strongly suggest that it will be capable of serving as a germ-line transformation vector in these and related species.

Recently Lozovskaya *et al.* (1996) successfully used *hobo* as a germ-line transformation vector in *D. virilis*, a species approximately 40 million years diverged from *D. melanogaster*. Although only two of the 398 germlines tested produced transgenic progeny, integration resulted from precise transpositional recombination with all of the hallmarks described above. The sequence of the 8 bp target site duplication closely resembled the consensus target reported by others. Although the *D. virilis* strain used in these experiments was not examined for the presence of endogenous *hobo* elements others have found no physical evidence for such elements in this species. Handler and Gomez (1995) have, however, reported the presence of a *hobo*-transposase-like activity which might originate from a related transposable element system.

The tephritid fruitfly, *B. tryoni* has recently been transformed using a *hobo* vector containing a gene conferring resistance to the antibiotic G418 as a selective marker (S. Whyard, D. O'Brochta, P. Atkinson, unpublished data). Transgenic G_1 insects were obtained at a moderate frequency (9.6%) and some of these integrations appear to have resulted from an interaction between the *hobo* vector and an endogenous *hobo*-like element (called *Homer*) known to exist in this species (see the following section). The nature of this interaction is currently under investigation but it has resulted in the inability to establish a stable transgenic line from these clearly transgenic individuals (S. Whyard, D. O'Brochta, P. Atkinson, unpublished data).

Related elements in other species

Atkinson *et al.* (1993) showed that *hobo* excision in *M. domestica* did not require a source of *hobo* transposase to recover *hobo* excision-reaction products. Because *hobo* excision in the host, *D. melanogaster*, was transposase-dependent Atkinson *et al.* (1993) proposed that *M. domestica* embryos must contain an endogenous source of *hobo* transposase-like activity. They went on to show that *M. domestica* contains a functional transposable element system (called *Hermes*) closely related to *hobo* (Warren *et al.*, 1994). Using a similar approach, the presence of a *hobo*-related systems have been detected in the drosophilids *D. virilis*, *D. melanica*, *D. repleta*, *D. saltans*, *D. willistoni* and *Chymomyza procnemis* (Handler and Gomez, 1995), the tephritids *Bactrocera tryoni* (S. Whyard, A. Pinkerton, D. O'Brochta and P. Atkinson, unpublished data), *B. curcubitae*, *B. dorsalis*, *C. capitata* and *Toxotrypana curvicauda* (A. Handler, personal communication) and the mosquitos *Aedes aegypti* and *Aedes australis* (A. Sarkar, K. Saville, K. Yardley, A. James, P. Atkinson and D. O'Brochta, unpublished data). Hence, using *hobo* excision as a bioassay for *hobo* or *hobo*-like transposase activity, a number of investigations have been able to detect the presence of *hobo*-like transposable element systems in species distantly related to *D. melanogaster* supporting the idea that *hAT* elements are widespread. The following is a brief description of *hAT* elements subsequently isolated from non-drosophilid insects.

Hermes (M. domestica). The presence of *Hermes* was originally inferred from genetic data obtained from *M. domestica* (described above) and isolated using the PCR with partially degenerate oligonucleotide primers homologous to conserved regions of *hobo*, *Ac* and *Tam3* transposases (Atkinson *et al.*, 1993). Warren *et al.* (1994) isolated and sequenced overlapping segments of several *Hermes* elements from *M. domestica* and constructed a 2749 bp consensus sequence. Full length *Hermes* elements contain a single long open reading frame (ORF) which, when conceptually translated, yields a protein of 612 amino acids that is 55% identical (71% similar) to *hobo* transposase (Fig. 2). *Hermes* is flanked by 17 bp imperfect terminal inverted repeats that are almost identical to those of *hobo's* and clearly related to the inverted terminal repeats of *Ac* and *Tam3* (Fig. 3). Comparison of *Hermes*, *hobo*, *Ac*, *Tam3* and an *Ac*-like element from pearl millet showed that the transposase of *Hermes* is most similar to that of *hobo* (Warren *et al.*, 1994). Limited efforts were made to assess *Hermes* sequence variation between *M. domestica* strains, and of eight strains examined, all showed evidence of at least one full-length or near full-length element. Six of the eight strains contained internally deleted *Hermes* elements while the remaining two appeared to contain only intact elements (Warren *et al.*, 1994). Therefore, *Hermes* is a transposable element that is closely related to *hobo* and based on structural criteria appeared functional. Direct tests of *Hermes* mobility, described in the next section, confirmed this hypothesis.

Hermit (L. cuprina)

hermit was isolated from the genome of the Australian sheep blowfly, *L. cuprina* (family, Calliphoridae) by screening a genomic DNA library with *hobo* sequences under conditions of reduced stringency (Coates *et al.*,

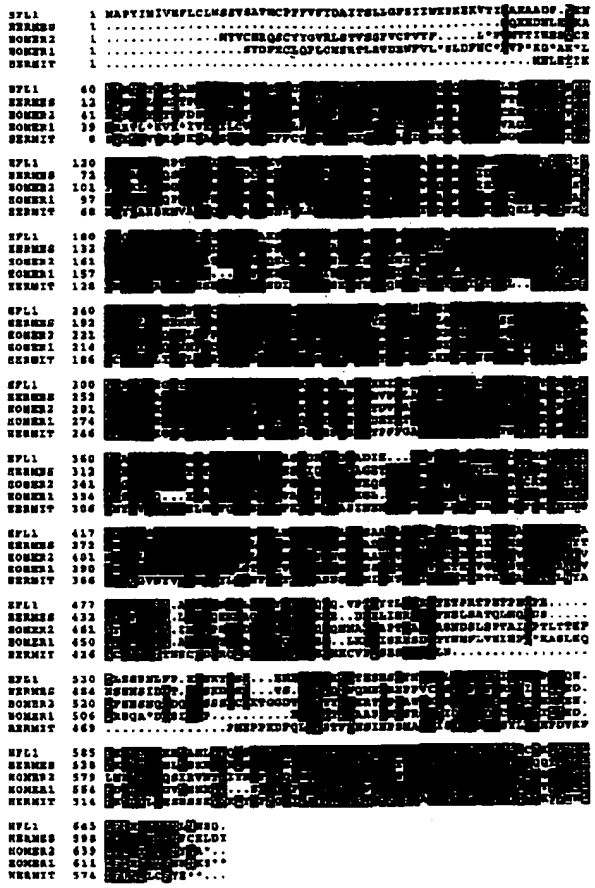


FIGURE 2. Alignment of the products from conceptual translation of the open reading frames of the (*hobo*), *Hermes*, *Homer 1*, *Homer 2* *HFLI* and *hermit* transposable elements showing the high levels of identity and similarity between them. Alignments were performed using the GCG software package from the University of Wisconsin, Madison, WI and displayed using BoxShade 3.0.

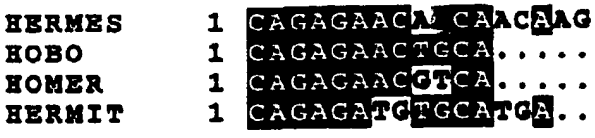


FIGURE 3. Alignment of *hAT* element inverted terminal repeat sequences showing their similarity. Note that the *Homer 1* sequence is provisional since the right hand inverted terminal repeat has yet to be identified.

1996). A single clone was recovered and DNA sequence analysis confirmed that it was a transposable element (*hermit*) 2716 bp in length with 15 bp terminal inverted repeats. Ten of the first 12 terminal nucleotides of *hermit* are identical to the corresponding nucleotides of the *hobo* element (Fig. 3). As with *Hermes* this similarity extended to other members of the *hAT* family. Southern blot analysis of genomic DNA indicated that *hermit* was present once in the genome of the strain of *L. cuprina* from which it was originally isolated and in 10 other strains from Australia, New Zealand and South Africa.

Hermit is 243 bp shorter than a full-length *hobo* element with most of this difference accounted for by two deletions within the putative ORF of the element. One deletion removes the N-terminal region of the protein while the second removes 18 amino acids from a region that is highly variable in all known *hAT*-element transposases. When comparing overall nucleotide identity *hermit* is 49%, identical to *hobo* and 51% identical to *Hermes*. Considering only the nucleotide sequence of the ORFs, *hermit* is 54% identical to *hobo* and 53% identical to *Hermes* (Fig. 2). Conceptual translation of the ORF within *hermit* results in a protein that is 64% similar and 42% identical to *hobo* transposase and 61% similar and 41% identical to the *Hermes* transposase.

Homer 1 and Homer 2 (B. tryoni)

As for the *Hermes* element of *M. domestica*, the presence of *hAT*-like elements in *B. tryoni* was originally inferred from *hobo* excision assays which showed that the *hobo* element could be recognized and mobilized by endogenous factors in developing embryos of *B. tryoni*. As with *Hermes* these *hAT* elements were isolated using a PCR-based approach. Unlike *Hermes*, *Hector* and *hermit*, however, there are at least two related forms of the *hAT* element present in *B. tryoni*, called *Homer 1* and *Homer 2* (A. Pinkerton, S. Whyard, D. O'Brochta and P. Atkinson, unpublished results). The presence of more than one form of transposable element in one species has been observed frequently for *mariner* elements but has not been previously recorded for *hAT* elements. *Homer 2* has an intact ORF suggesting that it encodes a functional transposase. The transposases of *Homer 1* and *Homer 2* are only 49% identical to each other but each is approximately 70% similar to *hobo* and *Hermes* transposases. Both elements are present in multiple copies within the genome of *B. tryoni*; there are approximately 5–10 copies of *Homer 1* and 20 copies of *Homer 2* per genome.

Hector (M. vetustissima)

Using a PCR-based strategy, similar to that employed to isolate *Hermes* from *M. domestica*, Warren *et al.* (1995) identified a sequence, called *Hector*, from the Australian bush fly, *M. vetustissima*, that is highly similar to *hobo* and *Hermes*. The *hAT* transposase-like sequence from *M. vetustissima*, when conceptually translated, encodes a protein 62% identical and 72% similar to the corresponding region of *hobo*. Although *M. vetustissima* and *M. domestica* are very closely related species *Hector* appears as similar to *hobo* as it is to *Hermes*. Surprisingly, *Hector* is a single copy sequence in the *M. vetustissima* strain from which it was isolated (Warren *et al.*, 1995).

Other insect hAT *elements*

Fragments of putative *hAT* elements have been isolated from other insect species. Using a PCR-based approach similar to that described above the existence of *hAT* elements in a number of tephritid species including the melon fly, *B. cucurbitae*, the oriental fruit fly, *B. dorsalis*, the Medfly, *C. capitata* and the Caribbean fruit fly, *Anas-*

trapha suspensa has been detected (A. Handler and S. Gomez, personal communication). Interestingly, over the short region of amino acid sequence available, the putative melon fly element is 74% identical and 86% similar to the corresponding region of the ORF of *Homer 2* from *B. tryoni*. Evidence for *hAT* elements in insect orders other than diptera is quite limited (DeVault and Narang, 1994) and likely reflects the limits of efforts to find these elements and not the limits of their distribution.

Mobility of non-drosophilid hAT *elements*

Direct evidence for mobility of *hAT* elements other than *hobo* is currently only available for *Hermes*. This element is transpositionally active in its host, *M. domestica* and in diverged non-host species (O'Brochta *et al.*, 1996). Using plasmid-based extrachromosomal transposition assays that have been so effective in assessing other short inverted repeat-type elements, the mobility of *Hermes* was assessed in *M. domestica*. *Hermes* accurately transposed in this species and the recombination products had all of the hallmarks of a genuine transposition event, including 8 bp insertion site duplications. The consensus integration site closely resembled that obtained for *hobo* elements (O'Brochta *et al.*, 1996). These data are significant because they demonstrate that *Hermes* is a functional transposable element from a non-drosophilid insect. Furthermore, these interplasmid transposition data were obtained from *M. domestica* that contain active endogenous *Hermes* elements. The ability of *Hermes* to transpose in such a strain suggests that, in *M. domestica*, any host-encoded molecules involved in repression or regulation of the element are insufficient to prevent transposition in these assays.

Hermes, like *hobo*, *Ac* and *Tam3* can also transpose in divergent species. The mobility of *Hermes* in non-host species is being tested using interplasmid transposition assays and by using it as a germ-line transformation vector. Using interplasmid transposition assays, accurate *Hermes* transposition has been detected in *L. cuprina* and *Cochliomyia macellaria* (family, Calliphoridae), *B. tryoni*, *Ceratitis capitata* (family, Tephritidae), *D. melanogaster* (family, Drosophilidae), *Ae. aegypti* (family, Culicidae) (A. Sarkar, C. Coates, J. Scura, K. Yardley, A. James, P. Atkinson and D. O'Brochta, unpublished results) and *Helicoverpa armigera* (order Lepidoptera, family Noctuidae) (Pinkerton *et al.*, 1996). *Hermes* transpositions in all cases created 8 bp insertion-site duplications within the target DNA, the consensus of which was similar to the target-site consensus sequence reported for *hobo*. In the case of *D. melanogaster* and *B. tryoni*, interplasmid transposition events were recovered at least 10 times more frequently than when similar experiments were performed in these species using *hobo*. In *Ae. aegypti* the rate of *Hermes* transposition was approximately 100-fold less than that observed for *Hermes* in *D. melanogaster*. *Hermes* appears to have a wide host range as well as elevated levels of activity when introduced into some non-host species. This property will make it particularly useful in deploying *Hermes* as a non-drosophilid gene vector.

Hermes has also been used as a germ-line transformation vector in *D. melanogaster* (O'Brochta *et al.*, 1996). Following the *P*-element paradigm, a binary vector system was constructed consisting of a non-autonomous *Hermes* element containing the *D. melanogaster* mini-*white* gene and a *Hermes* helper plasmid consisting of the *Hermes* transposase open reading frame under the control of the *D. melanogaster hsp70* promoter. Using standard protocols, a mixture of these plasmids was injected into preblastoderm embryos homozygous for a null mutation in *white*. The results confirmed that *Hermes* is an active element having mobility properties particularly useful for gene vector development. Of the 124 fertile G_0 adults recovered and individually tested, 43 (35%) yielded at least one transgenic, progeny. This transformation rate is within the range of rates usually encountered when using *P*-elements as a vector in this species. *Hermes* also showed evidence of being hyperactive in *Drosophila* (relative to *P*-elements). This was reflected in the high frequency of multiple integrations within single germlines. That is, 36 (84%) of the 43 transgenic G_0 germlines produced individuals containing two or more integrated *Hermes* elements. Furthermore, most (60%) of the G_0 transgenic germlines produced clusters of transgenic progeny. A cluster was a group of transgenic G_1 progeny comprising 10% or more of the total progeny arising from a single G_0 germline. In a small number of cases 90% or more of the progeny were transgenic indicating that almost every surviving germ cell following injection acquired an integrated *Hermes* element (O'Brochta *et al.*, 1996). The presence of clusters could reflect not only high rates of integration, but the integration of the transgene soon after being introduced into the embryo. In *D. melanogaster*, *Hermes* appears as active as any of the current gene vectors derived from *P*, *hobo* or *mariner* elements. Integrated *Hermes* elements were stable in the absence of *Hermes* transposase but remained capable of remobilization if *Hermes* transposase was reintroduced. All integrated *Hermes* elements examined were products of transpositional recombination involving only sequences precisely delimited by *Hermes* termini and resulted in the creation of characteristic 8 bp insertion site duplications. Little information is available regarding interactions between endogenous *hobo* elements and integrated *Hermes* elements since the transgenic *D. melanogaster* containing the *Hermes* vector was created in a strain lacking *hobo* elements. However, following the introduction of *hobo* transposase no *Hermes* revertants (excisions) were observed after screening 7228 chromosomes (M.M. Green, personal communication). When more sensitive plasmid-based excision assays are performed *Hermes* excision mediated by *hobo* transposase was observed at a rate of approximately 1 excision event per 5000 donor plasmids screened. (P. Sundararajan, P. Atkinson and D. O'Brochta, unpublished data). This is

approximately 10 fold less than the rate of *Hermes* excision observed in the presence of *Hermes* transposase in this species but it does provide evidence for interaction between these two elements. In conclusion, *Hermes* is one of the first functional non-drosophilid transposable elements to be isolated and harnessed as a gene vector. Based on its mobility properties it will certainly be useful as a germline transformation vector in some species and continued testing of the host range of existing *hAT* elements and the continued search for additional members of this family of elements in other insect taxa seem warranted.

THE *mariner* ELEMENT FAMILY

Description.

The *mariner* element was first isolated from the spontaneously arising *whitepeach* (*wpch*) mutant of *D. mauritiana* (Haymer and Marsh, 1986; Jacobson and Hartl, 1985). *mariner*(w^{pch}) is 1286 bp in length with a single long ORF, theoretically encoding a peptide of 346 amino acids. The element has 28 bp inverted terminal repeats containing four mismatches and creates a 2 bp TA repeat at its insertion site. *mariner* is present in 20–30 copies in the *D. mauritiana* genome but is absent from the sibling species, *D. melanogaster* (Jacobson *et al.*, 1986). The *mariner* element isolated from w^{pch}, despite the presence of an open reading frame does not encode a functional transposase. An autonomous *mariner* element, referred to as *Mos*, was identified and isolated from *D. mauritiana* and is capable of mobilizing other *mariner* elements *in trans*. *Mos* is identical to the *mariner* (w^{pch}) element except for six amino acid differences in the putative transposase coding region (Medhora *et al.*, 1988).

Mobility in the host

mariner elements can be highly mobile in *D. mauritiana* and have been used to generate mutations in the *white* and *yellow* genes of this species (Bryan *et al.*, 1990). When the autonomous *Mos* element was introduced into a strain containing many non-autonomous *mariner* elements visible mutations were recovered at a frequency of approximately 1 per 4000 X-chromosomes screened. Of seven mutant alleles recovered from this screen and examined at the molecular level, all contained an insertion of *mariner*. Mobility of *mariner* in its host, *D. mauritiana*, is not confined to the germline, unlike the movement of *P*- and *hobo* elements in *D. melanogaster*.

Mobility in non-hosts

The *mariner* and *Mos* elements are capable of transposition in *D. simulans* and *D. melanogaster*. The introduction of the *mariner*(w^{pch}) element into *D. simulans* by interspecific crosses between *D. mauritiana* and *D. simulans* resulted in the detection, isolation and characterization of autonomous *mariner* elements from *D. simulans* (Capy *et al.*, 1990; Maruyama and Hartl, 1991c). This "bioassay" approach for detecting functional *mariner* elements in other species is similar, in principle, to the bioassay approach that has been used for detecting functional *hobo*-like elements in non-drosophilid insects (Atkinson *et al.*, 1993; Handler and Gomez, 1995).

The mobility of the *mariner* element in non-host insects has been demonstrated in a variety of ways. For example, *mariner* has been used as a germline transformation vector of *D. melanogaster* following the *P*-element paradigm. Frequencies of *Mos*-mediated transformation varied from approximately 4 to 16% (Garza *et al.*, 1991; Lidholm *et al.*, 1993), however, these frequencies dropped considerably when additional DNA was inserted into the *mariner* element. When a 13.2 kb *mariner* element containing 11.9 kb of DNA containing the *white* gene (w^+) sequence was used as a transformation vector Lidholm *et al.* (1993) recovered only two w^+ transformants after testing 271 germlines. This rate of transformation is approximately 10-fold lower than that observed using *P*-elements under similar conditions. When the *mariner* vector was reconstructed with a mini-*white* gene of only 5.8 kb the transformation frequency did not improve, suggesting that *mariner* is relatively intolerant of the insertion of kilobases of additional sequences (Lohe *et al.*, 1995). Furthermore, the integrated *mariner* elements were somatically stable in *D. melanogaster*, despite subsequent introductions of various forms of the *Mos* element (expressing functional transposase) into these transgenic lines few progeny had evidence of somatic excision of the integrated *mariner* vector (Lidholm *et al.*, 1993; Lohe *et al.*, 1995). When attempts to remobilize the integrated *mariner*-w^+ element were made no evidence of transposition was reported. The dramatic reduction in transposition frequency of *mariner* element vectors containing kilobases of additional DNA together with their subsequent stability after integration has led to speculation that *mariner* vectors may have a strict limit to the amount of additional DNA they can carry (Lidholm *et al.*, 1993).

mariner is also capable of excision and transposition in non-drosophilid insects. Using plasmid-based assays (Fig. 1) Coates *et al.* (1995) recovered *mariner* excision-reaction products in the calliphorid, *L. cuprina*, and the tephritid *B. tryoni*. Excision-reaction products recovered from these species were identical in structure to those obtained using the same excision assay system in *D. melanogaster* and *D. mauritiana*. More recently these investigators used plasmid-based *mariner* transposition assays to assess *mariner* mobility in non-drosophilid insects. *mariner* was capable of accurate interplasmid transposition in developing embryos of both *D. melanogaster* and *L. cuprina* (Coates C., Turney C., Frommer M., O'Brochta D. and Atkinson P., unpublished results.) Only sequences precisely delimited by the terminal inverted repeats transposed and all integrations occurred into a TA dinucleotide target site. The frequency of transposition in *D. melanogaster* was similar to that seen using *hobo* and *Hermes* in this species under similar con-

ditions but the frequency of *mariner* transposition in *L. cuprina* was somewhat lower. These data demonstrate that *mariner* vectors containing up to 2.2 kb of additional DNA retain their mobility properties under these conditions suggesting they will be useful as gene vectors in insects if the amount of inserted DNA is minimized.

More recently the transposase protein from the *Himar 1* element (a *mariner* element from *Haematobia irritans*) was purified and shown to promote transpositional recombination *in vitro* in the absence of accessory proteins (Lampe D. and Robertson H., personal communication). These experiments elegantly demonstrate the autonomous nature of the *Himar 1* transposase and suggest that *Himar 1* will be capable of transposition in a wide variety of organisms. They also provide the opportunity to develop a germline transformation vector system which involves the direct injection of the *mariner* transposase protein together with the *Himar 1*-derived vector into developing insect embryonic germ cells. This will eliminate the need to transiently express the transposase gene in divergent insect hosts and may permit higher levels of transposase activity to be introduced into the embryo resulting in increased integration rates.

Related elements in other insects

Earlier efforts to determine the distribution of *mariner* elements revealed their presence in several drosophilid species, including *D. simulans, D. sechellia, D. teissieri, D. yakuba* and *Zaprionus tuberculatus* (Capy *et al.*, 1991; Maruyama and Hartl, 1991a, b). These elements were 97% identical and this uncommon homology among such divergent species together with the discontinuous phylogenetic distribution of these elements within this family of insects led to speculation that *mariner* has been transferred horizontally between species. The distribution of *mariner*-like elements was shown to be widespread after the serendipitous discovery by Lidholm *et al.* (1991) of a *mariner*-like element in the intron of the preprocecropin A gene of the Cecropia moth, *Hyalophora cecropia*. Using the sequence information from these highly diverged but related elements, Robertson (1993) used a PCR-based strategy to surveyed a large number of arthropod and non-arthropod species for the presence of additional *mariner*-family members. He discovered *mariner*-like elements in every major taxa in which he sampled, including insects. These data have recently been reviewed in some detail elsewhere (Robertson and Lampe, 1995). Based on this large data set numerous examples were found where the degree of sequence identity between *mariner*-like elements was far greater than expected if the elements were being transmitted solely by vertical transmission. Consequently, there appear to be numerous instances of horizontal transfer of *mariner*-like elements between highly divergent phylogentic groups. The mobility properties of *mariner* as determined experimentally and discussed above are consistent with the proposed ability of *mariner* to be horizontally transferred between species.

THE *Tc1* ELEMENT FAMILY

Description

The *Tc1* element is a short inverted repeat-type element originally identified and isolated from the nematode *Caenorhabditis elegans* (Emmons *et al.*, 1983). Related elements are found in a number of other organisms including insects. Four *Tc1*-like elements have been isolated from *Drosophila* species; *HB* and *Bari1* from *D. melanogaster* (Caizzi *et al.*, 1993; Harris *et al.*, 1988), *Uhu* from *D. heteroneura* (Brezinsky *et al.*, 1990) and *Minos* from *D. hydei* (Franz and Savakis, 1991). *Tc1*-like elements share similarities in their transposase coding regions and their inverted terminal repeats. *Tc1*-like elements also share limited similarities with *mariner* elements. Members of these two families of elements share approximately 10–20% amino acid identity, create TA dinucleotide repeats upon insertion into target DNA and may be derived from a common ancestral element. The reader is referred to the excellent review by Robertson (1995) for a comprehensive analysis of the similarities between *Tc1* and *mariner* elements. For the purposes of this mini-review and the construction of gene vectors the *Tc1*-like element from *D. hydei*, called *Minos*, is currently of the most interest.

Minos was originally identified as a dispersed repetitive sequence within the transcribed spacer of a rRNA gene of *D. hydei* (Franz and Savakis, 1991). *Minos* is 1.775 kb in length, has inverted terminal repeats of 255 bp and two long non-overlapping ORFs the longest of which shows 32% amino acid identity with the *Tc1* transposase. The two open reading frames are part of the same transcription unit and result in a spliced product. *Minos* is inserted at multiple locations within the genome of *D. hydei* and different strains of this species display insertion site polymorphisms (Franz *et al.*, 1994).

Mobility in host species

Tc1 is highly mobile in *C. elegans*, while *Minos* appears to be the only known active *Tc1*-like element from insects. *Minos* mobility was originally inferred from the presence of insertion site polymorphisms between very closely related strains of *D. hydei* (Franz *et al.*, 1994). Direct tests of *Minos* transposition in *D. hydei* have not been reported.

Mobility in non-host species

Minos is also mobile in non-host species of insects. Following the *P*-element paradigm a *Minos* vector containing the mini-*white* gene of *D. melanogaster* was constructed and injected into *D. melanogaster* embryos expressing *Minos* transposase (from a previously incorporated transgene) or into *D. melanogaster* embryos with a plasmid containing the *Minos* transposase gene under the regulatory control of the *D. melanogaster hsp* 70 promoter (Loukeris *et al.*, 1995a). Both strategies resulted in transformants at a frequency of approximately 2%. Four transgenic lines were examined for the presence of inte-

grated *Minos* elements and, in all cases, transposition of *Minos* into chromosomes was confirmed. Integrated *Minos* elements were delimited by the inverted terminal repeats and created TA dinucleotide duplications at the integration site (Loukeris *et al.*, 1995a). *Minos* therefore is a functional transformation vector for *D. melanogaster* although it is currently less efficient than *P-*, *hobo-* and *Hermes*-based vectors. Integrated *Minos* elements appear stable in the absence of *Minos*-encoded transposase despite the presence of related elements in this species.

Recently, *Minos* was successfully used as a germline transformation vector in the Medfly, *C. capitata*, a result that represents a major advance in the efforts to create technologies for the production of transgenic nondrosophilid insects (Loukeris *et al.*, 1995b). A *Minos* vector was constructed by replacing the transposase transcription unit with the cDNA of the *white* gene from the Medfly and under the regulatory control of the *hsp70* promoter of *D. melanogaster*. The *Minos* vector and transposase helper plasmids were co-injected into preblastoderm Medfly embryos homozygous for a recessive mutation in the *white* locus. 1.3% of the fertile G_0 adults developing from injected embryos produced at least one progeny with pigmented eyes from which lines were established. Vector integration was confirmed by in situ hybridization analysis of polytene chromosomes and Southern blot analysis of genomic DNA prepared from transformed individuals using *Minos* as a probe. The precise junction between the *Minos* vector and chromosomal DNA was not determined but, based on Southern blot analysis, integration resulted from transpositional recombination involving only the *Minos* vector. All transgenic lines analysed appeared to contain a single integrated *Minos* vector. The stability of the *Minos* vector within these transgenic insects is currently being investigated.

THE TTAA-SPECIFIC TRANSPOSABLE ELEMENT FAMILY

Description

TTAA-specific transposable elements are a diverse group of short inverted repeat-type elements that were identified as insertion sequences in baculoviruses that infect lepidoptera. These sequences were subsequently found within the genomes of the lepidoptera from which the baculoviruses were isolated (Beames and Summers, 1990; Carstens, 1987; Fraser *et al.*, 1985; Schetter *et al.*, 1990; Wang *et al.*, 1989). Information concerning the structure and biology of these elements is not extensive but they appear to be a diverse group of elements that share two features. First, the 5′ end of their terminal inverted repeats end in two or three C residues. Second, these elements are always found inserted in a TTAA target site. Beyond these similarities the elements have little in common. Their terminal inverted repeats are 13–15 nucleotides and their overall length can vary from 2.5 kb (*piggyBac* or IFP2) (Cary *et al.*, 1989; Fraser *et al.*, 1983) to 780 bp (*tagalong* or TFP2) (Fraser *et al.*, 1983). The *piggyBac* element, isolated originally as an insertion sequence in the *Galleria mellonella NPV* (*GmMNPV*) following passage in the *Trichoplusia ni*-derived cell line TN-368, encodes a single transcript of 2.1 kb in length (Fraser *et al.*, 1995). These elements currently show no similarities to other families of transposable elements and their distribution outside lepidoptera is unknown.

Mobility

TTAA-specific elements were discovered by virtue of their insertional mutagenic properties and consequently appear to be active transposable elements. None of the elements isolated, except the *piggyBac* (IFP2) element appears to be an autonomous element. *piggyBac* is mobile in cell lines derived from *T. ni* and *Spodoptera frugiperda*, and recent evidence demonstrates that it is capable of acting as a vector, i.e. transpose while carrying non-element DNA sequences. Using methods that resembled those described by O'Brochta *et al.* (1994), Fraser *et al.* (1995) assessed the ability of *piggyBac* to transpose in *S. frugiperda* cells. Fraser *et al.* (1995) constructed a *piggyBac* element marked with the β-galactosidase alpha peptide ORF and together with an intact *piggyBac* element and a wildtype baculovirus (AcMNPV) co-transfected *S. frugiperda* cells. Baculoviruses were subsequently recovered from infected cells that had the modified *piggyBac* element inserted into a non-essential region of the viral genome. Integration clearly resulted from transpositional recombination since only sequences delimited by the terminal inverted repeats of *piggyBac* transposed and all integration events created TTAA insertion-site duplications. Significantly transposition of *piggyBac* in these experiments required co-transfection of an intact *piggyBac* element suggesting that an element-encoded transposase was required. It is difficult to compare the rates of transposition in these experiments with those observed using *hobo* or *Hermes* because of the difference in experimental design. Nevertheless, these results are important because they provide evidence for a functional transposable element system from a family of lepidoptera (Noctuidae) of tremendous economic importance and represents a major advance in the efforts to develop non-drosophilid insect gene vectors. The components for a functional lepidopteran gene vector now appear to be available and tests of the practical utility of this system as measured by mobility *in vivo* will be forthcoming.

TTAA-specific elements appear to be present in a number of lepidopteran taxa and may be widespread within this order. For example, genetically marked *piggyBac* elements residing on baculoviruses or plasmids are capable of excising after introduction into lepidoptera cell lines that do not contain *piggyBac*-homologous sequences (M. Fraser, personal communication). Excision under these conditions resembles excision of *hobo* in *M. domestica* in the absence of *hobo*-encoded transposase (Atkinson *et al.*, 1993). There is a growing body of evidence indicating that cross-mobilization is due to the

presence of related and functional TTAA-specific elements in the species from which the cell lines originated.

LIMITS OF THE *P*-ELEMENT PARADIGM

Progress towards the development and use of transformation vectors using *mariner*, *Tc1*-like, TTAA and *hAT* elements affirms the utility of short inverted repeat-type elements and the general applicability of what we have referred to as the *P*-element paradigm to the problem of insect transformation. Despite the encouraging developments there will be limits or conditions under which the *P*-element paradigm will serve us poorly.

First, the *P*-element paradigm involves the use of specialized host strains of *Drosophila* as recipients of the transgene. *Drosophila* strains are used that do not possess full-length, functional copies of the transposable elements employed as transformation vectors in that species. The existence of strains devoid of *P*-, *hobo* and *mariner* elements has permitted the use of these elements as vectors in a genetic environment free of element-specific regulatory mechanisms that are responsible for repressing element movement. These element-specific repression systems can, however, be overcome under some conditions such as over-expressing functional transposase (Steller and Pirrotta, 1986). The presence of endogenous functional transposable elements of the type used as a vector may also compromise transgene stability and make maintenance of a stable transgenic line impossible. Although it is impossible to determine if strains completely devoid of elements from which the vector was constructed are absolutely required for the *P*-element paradigm to be effective in non-drosophilids, the presence of related elements may prove equally problematic. Strains of insects may be found that are devoid of the specific element being used as a gene vector but the presence of related elements may compromise the utility of the vector because of cross-mobilization. Cross-mobilization refers to the interaction of related but non-identical transposable elements and has been reported to occur between members of the *hAT* and TTAA families of elements. While strains of insects analogous to M (*P*-free) or E (*hobo*-free) strains of *D. melanogaster* would be desirable it may be possible to develop functional vector systems without them by designing "suicide systems" that permit the newly integrated vector to be disabled after integration. Vector disabling systems may be constructed that are analogous to the self-eviscerating systems used to remove antibiotic resistance genes from integrated plant gene vectors (Dale and Ow, 1991) and to create inducible "knockout" mice (Shastry, 1995). These systems depend upon the strategic placement of site-specific recombination signal sequences flanking the regions to be excised from the vector. Expression of the appropriate recombinase results in excision of the sequences between the recombination signals. By locating, for example, the recombination signal sequences for FLP recombinase, a highly active site-specific recombination system from yeast, in vector sequences essential for transposition and expressing FLP recombinase, the removal of essential vector sequences and permanent immobilization of the vector will result. It is premature to say if element-specific regulatory mechanisms will become a major obstacle to the development of transposable element-mediated transformation technology in non-drosophilid insects but its potential importance should be recognized.

The second feature of the *P*-element paradigm that may be difficult to satisfy in many insects and arthropods is the strategy employed to deliver the vector to the cells of interest. Within the *P*-element paradigm vector sequences are delivered directly to developing germcells by introducing DNA into the pole plasm of preblastoderm embryos. This has been usually done by injecting the vector but biolistics has also been used (Baldarelli and Lengyel, 1990). Unfortunately direct introduction of vector DNA into the pole plasm of preblastoderm embryos will not be possible for some insects because of unique qualities of the egg such as impenetrable chorions or because of the unique reproductive physiology of the insect. For example, viviparous insects (e.g. tsetse flies) will probably not be amenable to transformation following the *P*-element paradigm. Insects or arthropods that are minute or that have minute eggs may also be excluded from the strategies outlined in this review. Transforming these insect may require completely different strategies such as the use of viruses that can facilitate the delivery and entry of the DNA into a target cell and nucleus (Yee *et al.*, 1994). Direct injection of DNA into the gonads or hemocoel may be possible alternatives for transforming some species (Presnail and Hoy, 1992).

The *P*-element paradigm is clearly a powerful model upon which to base non-drosophilid insect transformation technology but its limits will be realized as it is applied to diverse insects. Other strategies will be required. Efforts to develop alternatives to the *P*-element paradigm are of great importance although not discussed in this review. The concept of a universal insect vector was never an entirely reasonable idea and even with the discovery of elements with broad host ranges it is unlikely that an element will be found that works in all species. Fortunately, there has been significant progress in the discovery and testing of transposable elements amenable to vector development and there are now multiple vector systems that can be anticipated (Table 1). It is now clear that the technical bottleneck created by the absence of non-drosophilid gene vectors has been breached and the widespread use of this technology by insect scientists will be forthcoming.

REFERENCES

Aarts M. G. M., Dirkse W. G., Stiekema W. J. and Pereira A. (1993) Transposon tagging of a male sterility gene *Arabidopsis*. *Nature* **363**, 715–717.

Anxolabehere D. and Periquet G. (1987) *P*-homologous sequences in

diptera are not restricted to the Drosophilidae family. *Genet. Iber.* **39**, 211–222.

Atkinson P. W., Warren W. D. and O'Brochta D. A. (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90**, 9693–9697.

Baker B., Schell J., Lorz H. and Federoff N. (1986) Transposition of the maize controlling element *Activator* in tobacco. *Proc. Natl. Acad. Sci. USA* **83**, 4844–4848.

Baldarelli R. M. and Lengyel J. A. (1990) Transient expression of DNA after ballistic introduction into *Drosophila* embryos. *Nucl. Acids Res.* **18**, 5903–5904.

Beames B. and Summers M. D. (1990) Sequence comparison of cellular and viral copies of host cell DNA insertions found in *Autographa Californica* nuclear polyhedrosis virus. *Virology* **174**, 354–363.

Beard C. B., O'Neill S. L., Mason P., Mandelco L., Woese C. R., Tesh R. B., Richards F. F. and Askoy S. (1993) Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Molec. Biol.* **1**, 123–131.

Bellen H. J., O'Kane C. J., Wilson C., Grossniklaus U., Pearson R. K. and Gehring W. J. (1989) *P*-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3**, 1288–1300.

Bingham P. M., Levis R. and Rubin G. M. (1981) The cloning of the DNA sequences from the *white* locus of *Drosophila melanogaster* using a novel and general method. *Cell* **25**, 693–704.

Blackman R. K. and Gelbart W. M. (1989) The transposable element *hobo* of *Drosophila melanogaster*. In *Mobile DNA* (Edited by Berg D. E. and How M. M.) pp. 523–529. American Society for Microbiology.

Blackman R. K., Macy M., Koehler D., Grimaila R. and Gelbart W. M. (1989) Identification of a fully-functional *hobo* transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J.* **8**, 211–217.

Brennan M. D., Rowan R. G. and Dickinson W. J. (1984) Introduction of a functional *P* element into the germ-line of *Drosophila hawaiiensis*. *Cell* **38**, 147–151.

Brezinsky L., Wang G. V., Humphreys T. and Hunt J. (1990) The transposable element *Uha* from Hawaiian *Drosophila*-member of the widely dispersed class of *Tc1*-like transposons. *Nucleic Acids Res.* **18**, 2053–2059.

Bryan G., Garza D. and Hartl D. (1990) Insertion and excision of the transposable element mariner in *Drosophila*. *Genetics* **125**, 103–114.

Caizzi R., Caggese, C. and Pimpinelli (1993) *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. *Genetics* **133**, 335–345.

Calvi B. R. and Gelbart W. M. (1994) The basis for germline specificity of the *hobo* transposable element in *Drosophila melanogaster*. *EMBO J.* **13**, 1636–1644.

Calvi B. R., Hong T. J., Findley S. D. and Gelbart W. M. (1991) Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*. *Cell* **66**, 465–471.

Capy P., Chakrani F., Lemeunier F., Hartl D. L. and David J. R. (1990) Active *mariner* transposable elements are widespread in natural populations of *Drosophila simulans*. *Proc. Royal Soc. Lond.* **242** 57–60.

Capy P., Maruyama K., David J. R. and Hartl D. L. (1991) Insertion sites of the transposable element *mariner* are fixed in the genome of *Drosophila sechellia*. *J. Mol. Evol.* **33**, 450–456.

Carstens E. B. (1987) Identification and nucleotide sequence of the regions of *Autographa californica* nuclear polyhedrosis virus genome carrying insertion elements derived from *Spodoptera frugiperda*. *Virology* **161**, 8–17.

Cary L. C., Goebel M., Corsaro B. G., Wang H. G., Rosen E. and Fraser M. J. (1989) Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**, 156–169.

Chuck G., Robbins T., Nijjar C., Ralston E., Courtney-Gutterson N. and Dooner H. K. (1993) Tagging and cloning of a Petunia flower color gene with the maize transposable element *Activator*. *Plant Cell* **5**, 371–378.

Clark J., Maddison W. and Kidwell M. (1994) Phylogenetic analysis supports horizontal transfer of *P* transposable elements. *Mol. Biol. Evol.* **11**, 40–50.

Coates C. J., Johnson K. N., Perkins H. D., Howells A. J., O'Brochta D. A. and Atkinson P. W. (1996) The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica* **97**, 23–31.

Coates C. J., Truney C. L., Frommer M., O'Brochta D. A., Warren W. D. and Atkinson P. W. (1995) *Mariner* can excise in non-drosophilid insects. *Molec. Gen. Genet.* **249**, 246–252.

Collins F. H. (1994) Prospects for malaria control through the genetic manipulation of its vectors. *Parasitol. Today* **10**, 370–371.

Curtis C. F. (1994) The case for malaia control by genetic manipulation of its vectors. *Parasitol. Today* **10**, 371–374.

Dale E. C. and Ow D. W. (1991) Gene transfer with subsequent removal of the selection gene from the host genome. *Proc. Natl. Acad. Sci. USA*, **88**, 10558–10562.

Daniels S. B., Chovnick A. and Boussy I. A. (1990) Distribution of *hobo* transposable elements in the genus *Drosophila*. *Mol. Biol. Evol.* **7**, 589–606.

Daniels S. B., Strausbaugh L. D., Ehrman L. and Armstrong R. (1984) Sequences homologous to *P* elements occur in *Drosophila paulistorum*, *Proc. Natl. Acad. Sci. USA*, **81**, 6794–6797.

DeVault J. D. and Narang S. K. (1994) Transposable elements in lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem. Biophys. Res. Commun.* **203**, 169–175.

Emmons S. W., Yesner L., Ruan K. S. and Katzenberg D. (1983) Evidence for a transposon in *Caenorhabditis elegans*. *Cell* **32**, 55–65.

Engels W. R. (1989) P elements in *Drosophila melanogaster*. In *Mobile DNA* (Edited by Berg D. E. and How M. M.) pp. 439–484. American Society for Microbiology.

Feldmar S. and Kunze R. (1991) The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain. *EMBO J.* **10**, 4003–4010.

Finnegan D. J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**, 103–107.

Franz G., Loukeris T. G., Dialektaki G., Thompson C. R. L., Savakis C. (1994) Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the *paired* DNA-binding domain, *Proc. Natl. Acad. Sci. USA*, **91**, 4746–4750.

Franz G. and Savakis C. (1991) *Minos*, A new transposable element from *Drosophila hydei*, is a member of the Tc1-like family of transposons. *Nucleic Acids Res.* **19**, 6646.

Fraser M. J., Brusca J. S., Smith G. E. and Summers M. D. (1985) Transposon-mediated mutagenesis of baculoviruses. *Virology* **145**, 356–361.

Fraser M. J., Cary L., Boonvisudhi K. and Wang H.-G. H. (1995) Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **211**, 397–407.

Fraser M. J., Smith G. E. and Summers M. D. (1983) The acquisition of host cell DNA sequences by baculoviruses: Relation between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* NPVs. *J. Virology* **47**, 287–300.

Garza D., Medhora M., Koga A. and Hartl D. L. (1991) Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster*. *Genetics* **128**, 303–310.

Hagemann S., Miller W. J. and Pinsker W. (1990) *P*-related sequences in *Drosophila bifasciata*: a molecular clue to the understanding of *P*-element evolution in the genus *Drosophila*. *J. Mol. Evol.* **31**, 478–484.

Handler A. M. and Gomez S. P. (1995) The *hobo* transposable element

has transposase-dependent and -independent excision activity in drosophild species. *Mol. Gen. Genet.* **247,** 399–408.

Harris L. J., Baillie D. L. and Rose A. M. (1988) Sequence identity between an inverted repeat family of transposable elements in *Drosophila* and *Caenorhabditis*. *Nucleic Acids Res.* **16,** 5991–5998.

Haymer D. S. and Marsh J. L. (1986) Germ line and somatic instability of a *white* mutation in *Drosophila mauritiana* due to a transposable genetic element. *Dev. Genet.* **6,** 281–291.

Hehl R. (1994) Transposon tagging in heterologus host plants. *Trends Genet.* **10,** 385–386.

Jacobson J. W. and Hartl D. L. (1985) Coupled instability of two X-linked genes in *Drosophila mauritiana*: Germinal and somatic mutability. *Genetics* **111,** 57–65.

Jacobson J. W., Medhora M. M. and Hartl D. L. (1986) Molecular structure of somatically unstable transposable element in *Drosophila*. *Proc. Natl. Acad. Sci. USA,* **83,** 8684–8688.

Kidwell M. (1992) Horizontal transfer of *P* elements and other short inverted repeat Tansposons. *Genetica* **86,** 275–286.

Kim A., Terzian C., Santamaria P., Pelisson A., Prud'homme N., Bucheton A. (1994) Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently in infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA,* **91,** 1285–1289.

Lansman R. A., Shade R. O., Grigliatti T. A., Brock H. W. (1987) Evolution of *P* transposable elements: sequences of *Drosophila nebulosa P* elements. *Proc. Natl. Acad. Sci. USA.* **84,** 6491–6495.

Lidholm D. A., Gudmundsson G. H. and Boman H. G. (1991) A highly repetitive, *mariner*-like element in the genome of *Hyalophora cecropia*. *J. Biol. Chem.* **266,** 11518–11521.

Lidholm D., Lohe A. and Hartl D. (1993) The transposable element mariner mediates germline transformation in drosophila melanogaster. *Genetics* **134,** 859–868.

Lim J. K., Simmons M. J., Raymond J. D., Cox N. M., Doll R. F., Culbert T. P. (1983) Homologue destabilization by a putative transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA,* **80,** 6624–6627.

Lohe A. R., Lidholm D. A. and Hartl D. L. (1995) Genotypic effects, maternal effects and grand-maternal effects of immobilized derivatives of the transposable element *mariner*. *Genetics* **140,** 183–192.

Loukeris T. G., Arca B., Livadaras I., Dialektaki G. and Savakis C. (1995a) Introduction of the transposable element *Minos* into the germ line of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **92,** 9485–9489.

Loukeris T. G., Livadaras I., Arca B., Zabalou S. and Savakis C. (1995b) Gene transfer into the Medfly, *Ceratitis capitata*, using a *Drosophila hydei* transposable element. *Science* **270,** 2002–2005.

Lozovskaya E. R., Nurminsky D. I., Hartl D. L. and Sullivan D. T. (1996) Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics* **142,** 173–177.

Maclean, N. (1995) (Ed.) *Animals with novel genes*. Cambridge: Cambridge University Press.

Maruyama K. and Hartl D. L. (1991a) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J. Molec. Evol.* **33,** 514–524.

Maruyama K. and Hartl D. L. (1991b) Evolution of the transposable element *mariner* in *Drosophila* species. *Genetics* **128,** 319–329.

Maruyama K. and Hartl D. L. (1991c) Interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. *J. Mol. Evol.* **33,** 514–524.

McCrane V., Carlson J. O., Miller B. R. and Beaty B. J. (1988) Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. *Am. J. Trop. Med. Hyg.* **39,** 502–510.

McGinnis W., Shermoen A. W. and Beckendorf S. K. (1983) A transposable element inserted just 5′ to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34,** 75–84.

Medhora M. M., MacPeek A. H. and Hartl D. L. (1988) Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *Mos* factor. *EMBO J.* **7,** 2185–2189.

Miller L. H., Sakai R. K., Romans P., Gwadz R. W., Kantoff P. and Coon H. G. (1987) Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science* **237,** 779–781.

Morris A. C., Eggelston P. and Crampton J. M. (1989) Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Med. Vet. Entomol.* **3,** 1–7.

O'Brochta D. A., Handler A. M. (1988) Mobility of *P* elements in drosophilids and non-drosophilids. *Proc. Natl. Acad. Sci. USA.* **85,** 6052–6056.

O'Brochta D. A., Warren W. D., Saville K. J. and Atkinson P. W. (1994) Interplasmid transposition of *Drosophila hobo* elements in non-drosophilid insects. *Mol. Gen. Genet.* **244,** 9–14.

O'Brochta D. A., Warren W. D., Saville K. J. and Atkinson P. W. (1996) *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics* **142,** 907–914.

Perkins H. D. and Howells A. J. (1992) Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina*. *Proc. Natl. Acad. Sci. USA.* **89,** 10753–10757.

Peterson P. W. and Yoder J. I. (1993) *Ac*-induced instability at the *Xanthophyllic* locus of tomato. *Genetics* **134,** 931–942.

Pinkerton A. C., O'Brochta D. A. and Atkinson P. W. (1996) Mobility of *hAT* transposable elements in the Old World American Bollworm, *Helicoverpa armigera*. Insect Mol. Biol., in press.

Presnail J. K., Hoy, M. A. (1992) Stable genetic transformation of a beneficial arthropod, *Metaseiulus occidentalis*. (Acai: Phytoseiidae), by a microinjection technique. *Proc. Natl. Acad. Sci. USA,* **89,** 7732–7736.

Rio D. C. (1991) Regulation of *Drosophila P* element transposition. *Trends Genet.* **7,** 282–287.

Robertson H. M. (1993) The *mariner* transposable element is widespread in insects. *Nature* **362,** 241–245.

Robertson H. M. (1995) The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiology* **41,** 99–105.

Robertson H. M. and Lampe D. J. (1995) Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **40,** 333–357.

Rubin G. M. and Spradling A. C. (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218,** 348–353.

Scavarda, N. J. and Hartl, D. L. (1984) Interspecific DNA transformation in *Drosophila*. *Proc. Natl. Acad. Sci. USA,* **81,** 7515–7519.

Schetter C., Oellig C. and Doerfler W. (1990) An insertion of insect cell DNA in the 81-map-unit segment of *Autographa californica* nuclear polyhedrosis virus DNA. *J. Virol.* **64,** 1844–1850.

Searles L. L., Jokerst R. S., Bingham P. M., Voelker R. A. and Greenleaf A. L. (1982) Molecular cloning of sequences from a *Drosophila* RNA polymerase II locus by *P* element transposon tagging. *Cell* **31,** 585–592.

Shastry B. S. (1995) Genetic knockouts in mice: an update. *Experientia* **51,** 1028–1039.

Simmons G. M. (1992) Horizontal transfer of *hobo* transposable elements within the *Drosophila melanogaster* species complex: evidence from DNA sequencing. *Mol. Biol. Evol.* **9,** 1050–1060.

Smith D., Wohlgemuth J., Calvi B. R., Franklin I. and Gelbart W. M. (1993) *Hobo* enhancer trapping mutagenesis in *Drosophila* reveals an insertion specificity different from *P* elements. *Genetics* **135,** 1063–1076.

Spielman A. (1994) Why entomological antimalaria research should not focus on transgenic mosquitoes. *Parasitol. Today* **10,** 374–376.

Spradling A. C. and Rubin G. M. (1982) Transposition of cloned *P* elements into *Drosophila* germ line chromosomes. *Science* **218,** 341–347.

Stamatis N., Yannopoulos G. and Pelecanos M. (1981) Comparative studies of two male recombination factors (MRF) isolated from a southern Greek *Drosophila melanogaster* population. *Genet. Res.* **38,** 125–135.

Steller H. and Pirrotta V. (1986) *P* transposons controlled by the heat shock promoter. *Molec. Cell Biol.* **6,** 1640–1649.

Streck R. D., MacGaffey J. E. and Beckendorf S. K. (1986) The struc-

ture of *hobo* transposable elements and their insertion sites. *EMBO J.* **5**, 3615–3623.

Walbot V. (1992) Strategies for mutagenesis and gene cloning using transposon tagging and T-DNA insertional mutagenesis. *Annu. Rev. Plant Phys. Plant Mol. Biol.* **43**, 49–82.

Walker V. K. (1989) Gene transfer in insects. *Adv. Cell Culture* **7**, 87–124.

Wang H. H., Fraser M. J. and Cary L. C. (1989) Transposon mutagenesis of baculoviruses: analysis of TFP3 lepidopteran transposon insertions at the FP locus of nuclear polyhedrosis viruses. *Gene* **81**, 97–108.

Warren W. D., Atkinson P. W. and O'Brochta D. A. (1994) The *Hermes* transposable element from the housefly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3*(*hAT*) element family. *Genetical Res. Camb.* **64**, 87–97.

Warren W. D., Atkinson P. W. and O'Brochta D. A. (1995) The Australian bushfly *Musca vetustissima*, contains a sequence related to transposons of the *hobo*, *Ac* and *Tam3* family. *Gene* **154**, 133–134.

Yee J. K., Friedmann T. and Burns J. C. (1994) Generation of high-titer pseudotyped retroviral vectors with very broad host range. *Methods Cell Biol.* **43**, 99–112.

Acknowledgements—The unpublished data provided by a number of colleagues was greatly appreciated. The comments of the reviewers were particularly helpful and also greatly appreciated. Work in the authors laboratories has been supported, in part, by the Common Wealth Scientific and Industrial Research Organization, the University of Maryland Biotechnology Institute, the U.S. Department of Agriculture and the National Institutes of Health.

ELSEVIER

Mutation Research 433 (1999) 159–168

DNA Repair

Review

Analysis of intrachromosomal homologous recombination in mammalian cell, using tandem repeat sequences

S. Lambert [1], Y. Saintigny [1], F. Delacote, F. Amiot, B. Chaput, M. Lecomte, S. Huck, P. Bertrand, B.S. Lopez *

UMR 217 CNRS, CEA, DSV, DRR, 60-68 Av. du Général Leclerc BP6 92 265 Fontenay aux Roses, Cedex, France

Received 6 January 1998; received in revised form 25 January 1999; accepted 10 February 1999

Abstract

In all the organisms, homologous recombination (HR) is involved in fundamental processes such as genome diversification and DNA repair. Several strategies can be devised to measure homologous recombination in mammalian cells. We present here the interest of using intrachromosomal tandem repeat sequences to measure HR in mammalian cells and we discuss the differences with the ectopic plasmids recombination. The present review focuses on the molecular mechanisms of HR between tandem repeats in mammalian cells. The possibility to use two different orientations of tandem repeats (direct or inverted repeats) in parallel constitutes also an advantage. While inverted repeats measure only events arising by strand exchange (gene conversion and crossing over), direct repeats monitor strand exchange events and also non-conservative processes such as single strand annealing or replication slippage. In yeast, these processes depend on different pathways, most of them also existing in mammalian cells. These data permit to devise substrates adapted to specific questions about HR in mammalian cells. The effect of substrate structures (heterologies, insertions/deletions, GT repeats, transcription) and consequences of DNA double strand breaks induced by ionizing radiation or endonuclease (especially the rare-cutting endonuclease ISce-I) on HR are discussed. Finally, transgenic mouse models using tandem repeats are briefly presented.

Keywords: Homologous recombination; Mammalian cells; Tandem repeat sequences

1. Introduction

Homologous recombination (HR) plays a central role in various fundamental processes determining genome organization and rearrangement such as molecular evolution [1,2] mating type switching in yeast (for review, see Ref. [3]), antigenic variation of the trypanosome (for review, see Ref. [4]), or for diversification of immunoglobulin genes in chicken [5] or in rabbit [6], faithful chromosomal segregation during meiosis in yeast (for review, Refs. [7,8]), and DNA repair (for review, see Ref. [9]).

On one hand, HR is involved in the maintenance of genome integrity by repairing damaged DNA; on the other hand, it may also contribute to genome instability since recombination between homologous

* Corresponding author. Tel.: +33-1-46-54-88-35; Fax: +33-1-46-54-91-80; E-mail: lopez@dsvidf.cea.fr

[1] Equal contribution.

0921-8777/99/$ - see front matter
PII: S0921-8777(99)00004-X

repeated sequences dispersed through the genome may lead to duplications, inversions, deletions, translocations [10]. Moreover, gene conversion with pseudogenes can inactivate a functional allele [11]. Finally, several lines of evidences connect HR with cancer predisposition or prevention. First, carcinogens stimulate intrachromosomal HR [12–15]. Second, intrachromosomal HR is elevated in DNA repair deficient cell lines, in Ataxia telangiectasia cell lines or p53 defective cells lines, all of these phenotypes being associated with cancer predisposition [16–20]. Third, Rad51, a putative recombination protein in mammalian cells, has been shown to interact with the products of the tumor suppressor genes *p53, BRCA1, BRCA2* [21–25].

The use of intrachromosomal tandem repeat represents a good strategy to measure HR in mammalian cells and has been often used because (i) it is the easiest way to introduce the two partners of recombination in a broad variety of cells and in one round of cell transfection, (ii) these systems permit efficient measure of HR and (iii) it allows to select for gene conversion or crossover and deletion events; (iv) depending on the orientation of the two markers, it can measure recombination arising by different mechanisms such as strand exchange (SE), single strand annealing (SSA) or replication slippage [26]. These mechanisms involving different pathways in yeast, it is thus essential to be able to distinguish between them in mammalian cells also. Here we present mechanisms of HR deduced from the knowledge in yeast, and also some molecular characteristics of HR between tandem repeat in mammalian cells. These data aid in the design of substrates adapted to more specific questions about the regulation of HR in mammalian cells.

1.1. Mechanisms and pathways of homologous recombination between tandem repeat: the knowledge from Saccharomyces cerevisiae

Two different orientations of the tandem repeat recombination substrate can be used: direct repeat (DR) or inverted repeat (IR). In bacteria and yeast, RecA/Rad51 protein promotes strand exchange (SE). SE recombination can act on DR as well as on IR and lead either to gene conversion or to crossing over (Fig. 1). Gene conversion between both DR and IR leaves intact the general structure of the locus. Depending on the orientation of the substrates, crossing over results in different structures: crossing over between IR leads to the inversion of the intervening sequence (Fig. 1B, Fig. 2A); crossing over between DR leads to the deletion of the intervening sequence either by an unequal sister chromatid exchange or by an intrachromatid exchange (Fig. 2B,C). In addition to SE recombination, direct repeats (DR) permit also to monitor deletion events arising by *RAD51* independent mechanisms such as single strand annealing (SSA) (Fig. 3), replication slippage and other mechanisms [26]. Because of their orientation, inverted repeats measure neither SSA nor replication slippage.

Extensive studies in yeast have determined genes involved in the control of HR. SE and SSA represent the main pathways using homologous sequences to achieve double strand break (DSB) repair. Besides DSB repair, replication slippage involved pathways different from those required for SE and SSA [26]. In *Saccharomyces cerevisiae*, the SE mechanism is dependent on the *RAD52* epistasis group including *RAD51, RAD52, RAD54, RAD55, RAD57* genes, whereas SSA does not involve these genes. SSA is dependent upon *SRS2*, on the nucleotide excision repair proteins Rad1 and Rad10 and on the mismatch repair proteins Msh2 and Msh3 [27,28]. Homologues to most of these genes have been described in mammalian cells (Table 1). Clearly, the study of HR in mammalian cells would benefit from the complementary use of the two kinds of substrates (DR and IR).

1.2. Molecular events of intrachromosomal recombination between duplicated sequences in mammalian cells

Two kinds of substrates have been used. One used two *LAC Z* sequences, recombinant cells being detected by the blue coloration. The second type of substrates used selectable genes (herpes simplex virus TK gene in tk^- cells, neomycin resistance gene, hygromycin resistance gene...), recombinant cells

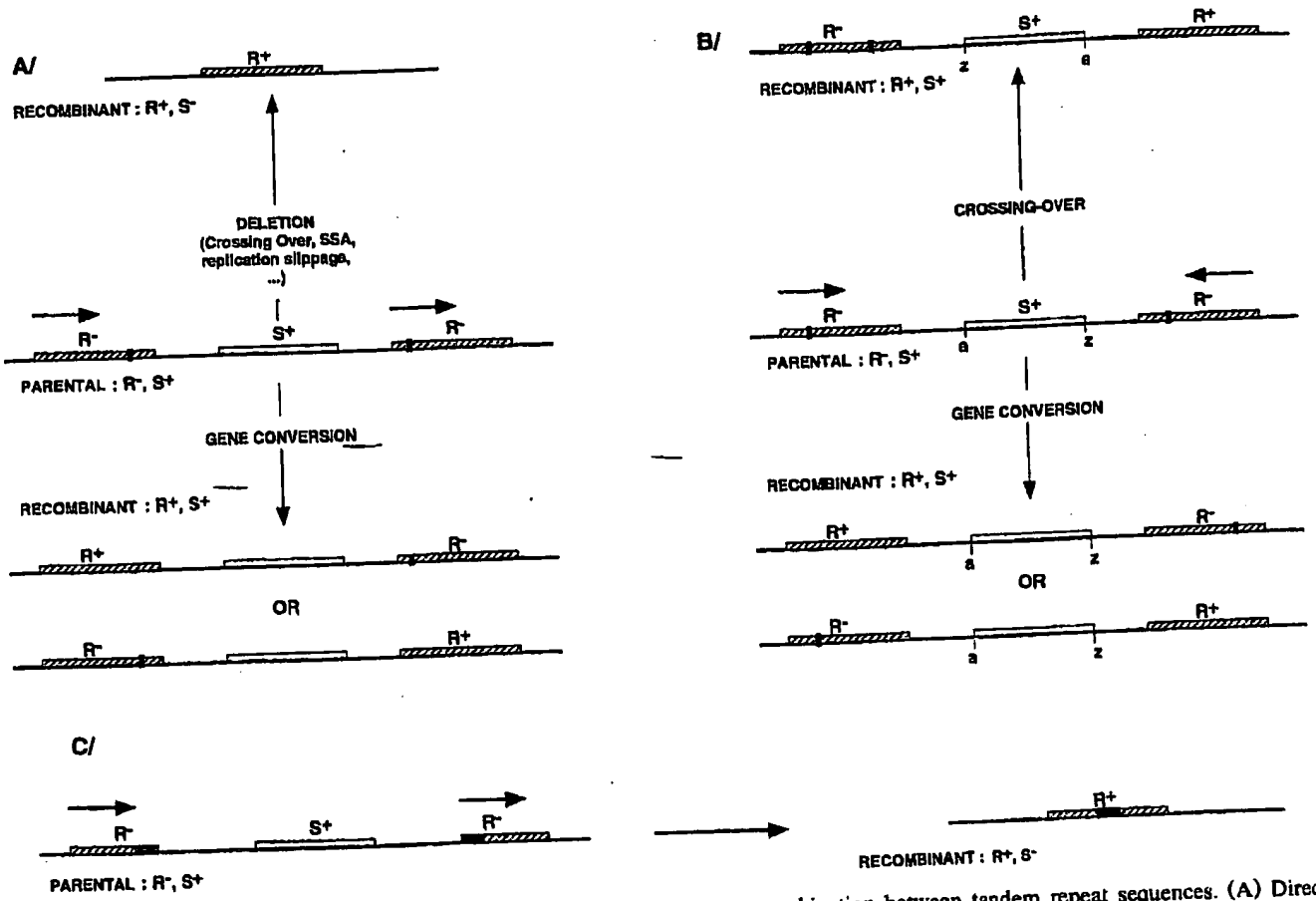


Fig. 1. Principle of the strategy to measure intrachromosomal homologous recombination between tandem repeat sequences. (A) Direct repeats, (B) inverted repeats (the letters a and z show the orientation of the intergenic sequence), (C) truncated repeats. The parental construction contains two inactive copies of a reporter gene (hatched boxes). This reporter gene can confer resistance to selective media. An alternative strategy uses the β-galactosidase gene whose expression can be monitored by histochemistry. In the parental form the two copies are inactivated (R^-) by a mutation, a deletion or an insertion (black boxes) located at different positions on each copy. An alternative strategy uses two truncated but overlapping regions (C) black boxes. In most of the systems, another selection gene (white boxes) is inserted between the two copies. The expression of the selection gene (S^+ or S^-) is used (i) to select the transfected cells with the recombination substrate, (ii) to analyze the distribution of recombination events (deletion or gene conversion). Recombination between the duplicated sequences creates a functional gene (R^+) either by gene conversion (a conservative event) or by deletion of the intergenic sequence (non-conservative event). Gene conversion can occur with direct as well as with inverted repeat sequences. Deletions can only occur with direct repeat sequences by crossing over after strand exchange (SE), by single strand annealing (SSA) or other events such as replication slippage. Crossovers between the inverted repeat sequences do not lead to deletion but to inversion of the intervening sequence. With two truncated copies, a deletion creates a functional gene (R^+).

forming clones when the selection was applied. Additionally, DSB repair has been studied following in vivo endonucleases treatment.

1.2.1. Spontaneous intrachromosomal recombination is a conservative mechanism

Analysis of the structure of recombinants has revealed that spontaneous intrachromosomal recombination is mainly a conservative mechanism in mouse L-cells with gene conversion representing approximately 80% of the recombination events [29]. This was observed using direct repeat or inverted repeat sequences [30]. The majority of the events giving rise to crossover products involved unequally paired of sister chromatids after DNA replication (see Fig. 2) [31]. These results contrast with those of extrachromosomal recombination which follow a non-conservative mechanism in mouse L-cells [32–

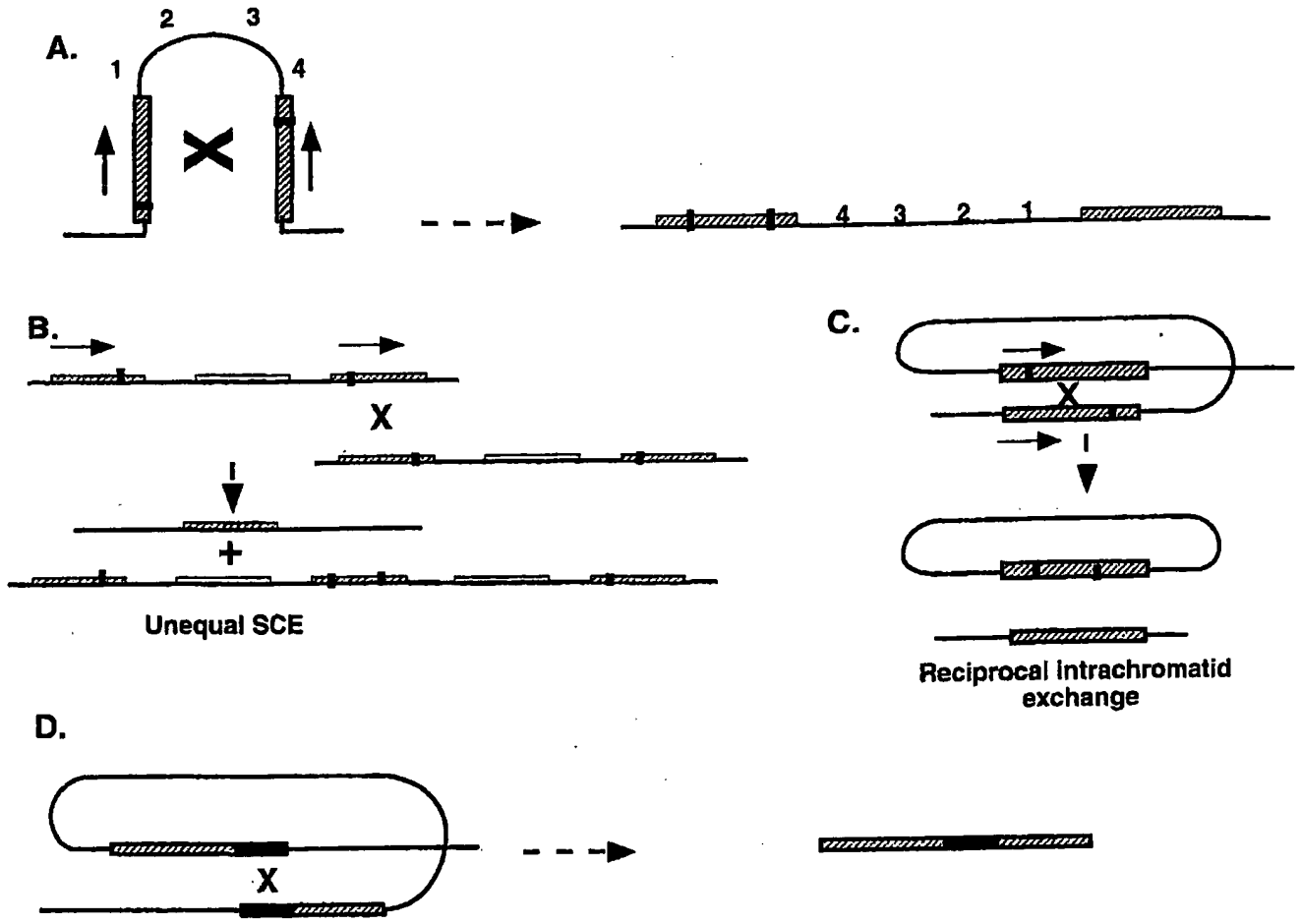


Fig. 2. Models for crossing-over events after SE. (A) Crossing over between two inverted repeats leads to the inversion of the intervening sequence. The arrows indicate the orientation of the two copies, (×) symbolizes the exchange and the numbers are here to orient the intergenic sequence. (B) and (C) recombination between two direct repeats (hatched boxes). The arrows indicate the orientation of the two copies. The lines represent the duplex DNA. (B) Recombination between two mispaired chromatids. (C) Intrachromatid recombination. The two copies (hatched boxes) are paired. After the exchange (×), the resolution of the Holliday junction can lead to gene conversion (see Fig. 1) or to a crossing over event. In (B), the crossing over leads to the formation of a unique active gene and to the deletion of the intergenic sequence on one chromatid and to a triplication on the other chromatid. In (C), the crossing over leads to the maintenance of only one copy on the chromosome and one loop that can be eliminated. In some cases, the loop can be integrated elsewhere in the genome. (D) Crossover between two truncated copies. The black box represents the overlapping sequence. (A, C, D) correspond to reciprocal intrachromatid exchanges.

34]. It cannot be excluded that these ratios may differ in other cell types.

1.2.2. Heteroduplex formation

The different models for HR involved intermediate structures in which DNA-strand exchanges create hybrids and heteroduplex DNA between the two recombining molecules [35–37]. Biochemical studies of mitotically dividing mammalian cells indicate that exchange between sister chromatids involved interstrand transfer of DNA [38,39]. In addition, heteroduplex formation during HR promoted by human nuclear extracts was observed in a cell free system [40]. The analysis of the intrachromosomal recombination products between tandem repeats has shown that recombinant colonies result from unrepaired heteroduplex DNA [31,41]. Since particular structures may affect the formation, the stability or the resolu-

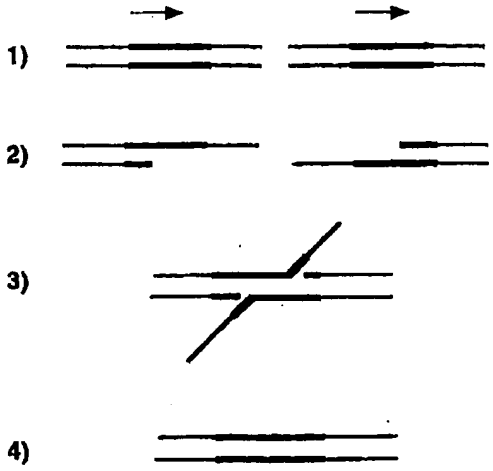


Fig. 3. The single strand annealing model [32]. The two complementary strands of the duplex DNA are drawn here. The heavy lines symbolized the homologous repeat sequences and the arrows the orientation of the two repeats. (1) A double strand break can occur even in the non-homologous intervening sequence. (2) After degradation by a single strand exonuclease, single stranded tails (ss-tails) are created. If the repeats are in a direct orientation, complementary ss-tails are created. (3) Annealing of the two complementary single stranded regions results in a structure leading to the deletion of the intervening sequence after the resolution of this structure. (4) SSA is exclusively a non conservative mechanism. If the repeats are in an inverted orientation, the ss-tails (step 2) are identical but not complementary and SSA cannot act.

tion of the heteroduplex recombination intermediates, the effect on recombination of the DNA structure (insertions or deletions leading to loops formation into the heteroduplex molecule the homology requirements...) have been studied.

1.2.3. Effect of the sequence structure: insertions, deletions, sequence homology requirements, microsatellites, transcription

1.2.3.1. Insertions and deletions. The molecular nature of insertion or deletion mutations (corresponding to the black boxes in Figs. 1 and 2) in the copies of the duplication can influence the efficiency of HR. DNA strand exchange is able to propagate through heterologous sequences forming heteroduplex DNA bearing loops. The resolution of such intermediates leads more frequently to the excision of the loop with an efficiency correlated to the size of the loop [42].

1.2.3.2. Homology requirements. Two alternative strategies can be drawn to address the question of how much homology is required for HR in mammalian cells. The first uses two truncated molecules with different lengths of overlapping of uninterrupted homology (Fig. 2D). When the two interacting molecules share length of homologies between 295 bp and 1.8 kb, the rate of gene conversion is

Table 1
Comparison between strand exchange (SE) and single strand annealing (SSA)

Process	Products	Orientation of the substrates	Pathway in *S. cerevisiae*	Homologues in human cells
Strand exchange	Gene conversion	Direct repeat	*RAD51*[a]	*HsRAD51*[a]
		Inverted repeat	*RAD52*[b]	*HsRAD52*
			RAD54	*HsRAD54*
	Crossing over		*RAD55*[a]	see:[a]
	Deletions	Direct repeat	*RAD57*[a]	see:[a]
	Inversions	Inverted repeat	*RAD59*	
Single strand annealing	Deletions	Direct repeat	*RAD1*	*XP-F*[c]
			RAD10	*ERCC1*
			MSH2	*HsMSH2*
			MSH3	*HsMSH3*
			SRS2	

[a] In yeast *RAD55* and *RAD57* share homologies with *RAD51*. In mammalian cells, beside *HsRAD51*, at least 6 other *RAD51* homologues have been described: *XRCC2*, *XRCC3*, *RAD51B/HREC2*, *RAD51H3*, *RAD51C*, *RAD51D*. Some of these homologues presumably correspond to *RAD55* and to *RAD57*.

[b] Defines the epistasis group for homologous recombination in yeast.

[c] *Xeroderma pigmentosum* group F.

directly proportional to the length of uninterrupted homology. This rate is reduced 7 fold with 200 bp homology and 100 fold with 95 bp length of homology [43].

The second strategy uses two molecules of approximately the same size but containing sequence polymorphism. Waldman and Liskay used a HSV-*tk* duplication in which one *TK* copy came from HSV type 1 and the other copy from HSV type2 (homeologous recombination). These *TK* genes share 81% of homology. The authors observed that with 19% divergence, the rate of intrachromosomal HR was reduced 1000 fold relative to the rate of HR between two identical HSV1-*tk* sequences. In contrast, the rate of intramolecular or intermolecular extrachromosomal recombination was only reduced by a factor 3 to 15 [44]. These results also argue in favor of distinct mechanisms between extrachromosomal and intrachromosomal recombination. Moreover, if efficient intrachromosomal recombination required a minimum of 134 to 232 bp of uninterrupted homology, a single-nucleotide heterology in this minimal region of homology was sufficient to inhibit efficient recombination [45]. In addition, when recombination initiates in a perfectly homologous sequence, it is able in a second step to propagate through an adjacent sequences exhibiting 19% heterologies [45]. Finally, Yang and Waldman [46] show that gene conversion involved transfer of uninterrupted blocks of information from 35 to more than 330 bp. Taken together, these results are consistent with the notion that more than 200 bp of homology are required to initiate efficient gene conversion in mammalian cells. This has also been found in in vitro reactions in cell free systems [47].

1.2.3.3. Effect of GT repeats. A variety of DNA sequences may play direct or indirect roles in recombination by their effects on the DNA structures. It was proposed that GT and GC repeats, which can form Z-DNA, may influence recombination [48]. It has been shown that GT, GC repeats and minisatellite repeats, stimulated extrachromosomal recombination of transfected DNA [49,50]. However, using an intrachromosomal assay, a $(GT)_{29}$ repeat has been shown to be unable to stimulate recombination and to influence the distribution of the recombination events [51].

1.2.3.4. Transcription. Transcription stimulates homologous recombination in *S. cerevisiae* [52,53]. Stimulation by transcription of extrachromosomal recombination [54] and intrachromosomal recombination [55] has been reported in CHO cells. Alleles transcribed at high levels recombined about 2 to 7-fold more frequently than identical alleles transcribed at low level. In line with this, preferential repair of UV damage has been shown to abolish the transcription–stimulation of HR [56]. Finally, transcription has no effect on recombination induced by a DNA double strand break [57].

1.3. Stimulation of homologous recombination by DNA double strand breaks

In different species, DNA-damaging agents were shown to stimulate homologous recombination. DNA double strand breaks (DSB) are the main lesion involved in the stimulation of homologous recombination in different organisms. It is used to initiate recombination during meiosis in yeast *S. cerevisiae* (for review, see Refs. [7,8]). The DSB is also one of the most genotoxic lesion induced by ionizing radiation and can be repaired by two general mechanisms: (i) non homologous end joining (NHEJ), which does not necessarily involve sequence homologies and promote the end-joining of the broken extremities; (ii) homology-directed repair which involves homologous sequences and corresponds to SE and SSA mechanisms. Ionizing radiation stimulates interallelic recombination in the endogenous *TK* locus in a human lymphoblast cell line [58]. However, the effect of ionizing radiation on intrachromosomal HR between tandem repeat sequences varies according to the cells and/or to the substrates used. Although, HR between two LacZ sequences in CV-1 cells is stimulated by ionizing radiation [59], it does not stimulate HR between HSV-TK sequences in mouse L-cells [12]. The fact that the cell lines used are different could explain this result. Another explanation could involved the differences in the substrates, LacZ vs. TK sequences. In the former case, recombinant are scored by coloration (β-galactosidase activity) some hours following the treatment. In contrast TK^+ recombination events can be scored on surviving colonies several weeks after the treatment. Thus,

in the former system, recombination events can be scored before the death of the cells generally arising several days after exposure to radiation. In contrast, the second system scores only surviving colonies.

Another way to produce DSB in target DNA uses nucleases. Restriction enzymes corresponding to single sites present in the target were electroporated into CHO cells or human cells [42,60]. In these cases, recombination was increased by more than 10-fold. Electroporation of the rare cutting yeast endonuclease PI-*SceI* also stimulated recombination provided there was a corresponding cleavage site in the duplication [60]. The yeast endonuclease I-SceI has provided a useful tool to study targeted DSB in mammalian cells, as already reviewed [61]. I-SceI recognizes a cleavage site of 18 bp long. Due to this large restriction sequence, there is probably no I-SceI site in the mammalian genome and expression of the I-SceI enzyme in mammalian cells is not toxic. However, when the I-SceI restriction site is present in the duplication, at least 80% of the molecules are cleaved after transfection of a plasmid expressing I-SceI enzyme [62]. Induction of a site-directed DSB into the recombination substrates strongly stimulates both homologous and non-homologous recombination [63,64,57]. HR is stimulated 100 times but non-homologous recombination is stimulated 1000 times [64]. However, using a physical analysis, homology-directed repair of I-SceI-induced DSB's is found to account for 30–50% of the observed events [65]. ISce-I induced recombination produced mainly deletion events (80%), that are interpreted as a result from SSA. Finally, in contrast with spontaneous HR, transcription does not stimulate DSB-induced recombination [57], although we cannot exclude that the efficiency of DSB-stimulation is so great that it would mask the effect of transcription.

1.4. Transgenic mice models

The assay using duplicated repeats to measure intrachromosomal recombination can be envisioned in vivo in transgenic mice. In that case, recombination can only be recorded using genes giving a coloration such as the β-galactosidase gene, but not with a gene conferring resistance to a drug. Two models were developed to measure recombination in specific tissues. One model has been developed to analyze the lineage of cells in the myotome; thus, the expression of the recombination substrate was driven by a promoter which confers expression specifically to cells of this compartment: the promoter of the α-subunit of acetylcholine receptor. The descendants of the recombinant cells are histochemically identified, permitting the analysis of the lineage of cells in the intact embryos [66]. It appears that the frequency of recombination in this tissue is similar to the frequency measured in cultured cells: between 1 and 2×10^{-6} [66].

Another transgenic model also uses a duplication of β-galactosidase genes driven by a meiosis specific promoter. Germ line gene conversion was analyzed in transgenic male gametes. Spermatids which undergo intrachromosomal gene conversion produce functional β-galactosidase ($lacZ^+$), visualized by histochemical staining. Approximately 2% of the spermatids, produced by a combination of meiotic and mitotic conversion events, were $LacZ^+$ [67].

2. Conclusions

Numerous differences exist between intrachromosomal and extrachromosomal HR. For example, intrachromosomic recombination is conservative whilst plasmid recombination is non-conservative in mouse L-cells; in addition heterologies and GT repeats differently affected plasmid and intrachromosomal recombination. Moreover, the number of plasmid copies is not controlled and it is well established that plasmids are submitted to extensive nuclease attack after transfection. Chromatin structure and the nuclear organization may also account for the differences observed between plasmid and intrachromosomal recombination. Thus, we believe it is most suitable to measure HR between two intrachromosomal tandem repeat sequences.

The other possibilities to measure HR between two intrachromosomal sequences are interallelic and ectopic recombination, which are however technically complicate to set up. Furthermore, interallelic HR as well as ectopic HR are rather inefficient in mammalian cells [68,42,69]. Finally, the possibility of using two different orientations of tandem repeats

(direct or inverted) and the use of the rare-cutting endonuclease ISce-I provide a set of complementary approaches to study HR in mammalian cells.

Using these of strategies, connections between HR and other fundamental cellular processes have been described. Firstly, transcription stimulates HR but not DSB-induced HR [57]. Connections with cell cycle control and DNA repair have also been studied: alteration of *ATM* or p53 functions lead to an increase of HR [18–20]; carcinogens and DNA damaging agents treatments stimulate both gene conversion and deletion [12]. More generally, the efficiency of repair of DNA damages diminishes the stimulation of recombination [17,56]; finally, inhibition of Poly(ADP-ribose)polymerase (PARP) by 3-methoxybenzamide increases recombination between 3 and 4-fold in mouse L-cells [70].

Very little is known on the genetic control of homologous recombination in mammalian cells. Overexpression of human *RAD52* stimulates recombination between two lacZ direct repeats about 3-fold, in monkey cells and confers resistance to ionizing radiation [59]. The Rad52 stimulation acts via the single stranded DNA binding protein RpA which is also involved in replication and DNA excision repair [71]. One question would be to understand how overexpression of only one component of a multiprotein complex can stimulate the whole process, i.e., homologous recombination. One explanation could be that Rad52 is the limiting factor of the complex. However, it has recently been shown that overexpression of the Hamster Rad51 protein into CHO, also confers resistance to ionizing radiation and stimulates HR between direct repeat sequences [72]; this result is not in accordance with this hypothesis. The combined analysis of HR between direct repeats, between inverted repeats and of the DSB-induced HR should afford essential information to understand these process in mammalian cells.

Acknowledgements

This work was supported by CEA, CNRS, ARC (1366) and Electricité de France. We thank C. White and C. Ducray for helpful comments.

References

[1] S.A. Liebhaber, M. Goosens, Y.-K. Kan, Homology and concerted evolution at the α 1 and α 2 loci of human α-globin, Nature 290 (1981) 26–29.
[2] D. Baltimore, Gene conversion: some implications for immunoglobulin Genes, Cell 26 (1981) 295–296.
[3] J.E. Haber, Mating type gene switching in *Saccharomyces cerevisiae*, TIG 8 (1992) 446–452.
[4] L.H.T. Van der Ploeg, Antigenic variation in African trypanosome, TIG 8 (1992) 452–457.
[5] C.A. Reynaud, V. Anquez, H. Grimal, J.-C. Weill, A hyperconversion mechanism generates the chicken light chain preimmune repertoire, Cell 48 (1987) 379–388.
[6] R.S. Becker, K.L. Knight, Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits, Cell 63 (1990) 987–997.
[7] S.G. Roeder, Chromosome synapsis and genetic recombination, TIG 6 (1990) 385–389.
[8] N. Kleckner, Meiosis: how could it work?, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 8167–8174.
[9] E.C. Friedberg, G.C. Walker, W. Siede, DNA Repair and Mutagenesis, ASM Press, Washington, DC, 1995.
[10] R.J. Bollag, A.S. Waldman, R.M. Liskay, Homologous recombination in mammalian cells, Annu. Rev. Genet. 23 (1989) 199–225.
[11] M. Amor, L.K. Parker, H. Globerman, M.L. New, P.C. White, Mutation in the CYP21B gene (Ile-172-Asn) causes steroid 21-hydroxylase deficiency, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 1600–1604.
[12] Y. Wang, V.M. Maher, R.M. Liskay, J.J. Mc-Cormick, Carcinogens can induce homologous recombination between duplicated chromosomal sequences in mouse L cells, Mol. Cell. Biol. 8 (1988) 196–202.
[13] D. Hellgren, Mutagen-induced recombination in mammalian cells in vitro, Mut. Res. 284 (1992) 37–51.
[14] D.L. Ludwig, J.R. Stringer, Spontaneous and induced homologous recombination between lacZ chromosomal direct repeats in CV-1 cells, Somat. Cell. Mol. Genet. 20 (1994) 11–25.
[15] H. Zhang, T. Tsujimura, N.P. Bhattacharyya, V.M. Maher, J.J. McCormick, O6-methylguanine induces intrachromosomal homologous recombination in human cells, Carcinogenesis 17 (1996) 2229–2235.
[16] T. Tsujimura, V.M. Maher, A.R. Godwin, R.M. Liskay, J.J. McCormick, Frequency of intrachromosomal homologous recombination induced by UV radiation in normally repairing and excision repair-deficient human cells, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 1566–1570.
[17] N.P. Bhattacharyya, V.M. Maher, J.J. McCormick, Effect of nucleotide excision repair in human cells on intrachromosomal homologous recombination induced by UV and 1-nitrosopyrene, Mol. Cell. Biol. 10 (1990) 3945–3951.
[18] S.M. Meyn, High spontaneous intrachromosomal recombination rates in ataxia–telangiectasia, Science 260 (1993) 1327–1330.

[19] P. Bertrand, D. Rouillard, A. Boulet, C. Levalois, T. Soussi, B.S. Lopez, Increase of spontaneous intrachromosomal homologous recombination in mammalian cells expressing a mutant p53 protein, Oncogene 14 (1997) 117–1122.

[20] K.L. Mekeel, W. Tang, L.A. Kachnic, C.-M. Luo, J.S. DeFrank, S.N. Powell, Inactivation of p53 results in high rates of homologous recombination, Oncogene 14 (1997) 1847–1857.

[21] H.-W. Sturzbecher, B. Donzelmann, W. Henning, U. Knippschild, S. Buchhop, p53 is linked directly to homologous recombination processes via RAD51/RecA protein interaction, EMBO J. 15 (1996) 1992–2002.

[22] S. Buchhop, M.K. Gibson, X.W. Wang, P. Wagner, H.-W. Sturzbecher, C.C. Harris, Interaction of p53 with the human Rad51 protein, Nucleic Acids Res. 25 (1997) 3868–3874.

[23] R. Scully, J. Chen, A. Plug, Y. Xiao, D. Weaver, J. Feunteun, T. Ashley, D.M. Livingston, Association of BRCA1 with Rad51 in mitotic and meiotic cells, Cell 88 (1997) 265–275.

[24] R. Mizuta, J.M. LaSalle, H.-L. Cheng, A. Shinohara, H. Ogawa, N. Copeland, N.A. Jenkins, M. Lalande, F.W. Alt, RAB22 and RAB163/mouse BRCA2: proteins that specifically interact with the RAD51 protein, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 6927–6932.

[25] L.Y. Marmorstein, T. Ouchi, S.A. Aaronson, The BRCA2 gene product functionally interacts with p53 and RAD51, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 13869–13874.

[26] H.L. Klein, Genetic control of intrachromosomal recombination, BioEssays 17 (1995) 147–159.

[27] E.L. Ivanov, N. Sugawara, J. Fishman-Lobell, J.E. Haber, Genetic requirements for the single-strand annealing pathway of double-strand break repair in *Saccharomyces cerevisiae*, Genetics 142 (1996) 693–704.

[28] N. Sugawara, F. Pâques, M. Colaiàcovo, J.E. Haber, Role of Msh2 and Msh3 repair proteins in double-strand break induced recombination, Proc. Natl. Acad. Sci. U.S.A. 94 (1997) 9214–9219.

[29] R.M. Liskay, J.L. Stachelek, A. Letsou, Homologous recombination between repeated chromosomal sequences in mouse cells, Cold Spring Harbor Symp. Quant. Biol. 49 (1984) 183–189.

[30] R.J. Bollag, R.M. Liskay, Conservative intrachromosomal recombination between inverted repeats in mouse cells: association between reciprocal exchange and gene conversion, Genetics 119 (1988) 161–169.

[31] A. Letsou, R.M. Liskay, Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells, Genetics 117 (1987) 759–769.

[32] F.L. Lin, K. Sperle, N. Sternberg, Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process, Mol. Cell. Biol. 4 (1984) 1020–1034.

[33] S. Chakrabarti, M.M. Seidman, Intramolecular recombination between transfected sequences in mammalian cells is nonconservative, Mol. Cell. Biol. 6 (1986) 2520–2526.

[34] F. Lin, K. Sperle, N. Sternberg, Extrachromosomal recombination in mammalian cells as studied with single- and double stranded DNA substrates, Mol. Cell. Biol. 7 (1987) 129–140.

[35] R. Holliday, A mechanism for gene conversion in fungi, Genet. Res. 5 (1964) 282–306.

[36] M.S. Meselson, C.M. Radding, A general model for genetic recombination, Proc. Natl. Acad. Sci. U.S.A. 72 (1975) 358–361.

[37] J.W. Szostak, T.L. Orr-Weaver, R.J. Rothstein, F. Stahl, The double-strand-break repair model for recombination, Cell 33 (1983) 25–35.

[38] J. Rommelaere, A. Miller-Faures, Detection by density equilibrium centrifugation of recombinant-like DNA molecules in somatic mammalian cells, J. Mol. Biol. 98 (1975) 195–218.

[39] P.D. Moore, R. Holliday, Evidence for the formation of hybrid DNA during mitotic recombination in Chinese hamster cells, Cell 8 (1976) 573–579.

[40] B. Lopez, S. Rousset, J. Coppey, Homologous recombination intermediates between two duplex DNA catalysed by human cell extracts, Nucleic Acids Res. 15 (1987) 5643–5655.

[41] R.J. Bollag, D.R. Elwood, E.D. Tobin, A.R. Godwin, R.M. Liskay, Formation of heteroduplex DNA during mammalian intrachromosomal gene conversion, Mol. Cell. Biol. 12 (1992) 1546–1552.

[42] A.R. Godwin, R.J. Bollag, D.M. Christie, R.M. Liskay, Spontaneous and restriction enzyme-induced chromosomal recombination in mammalian cells, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 12554–12558.

[43] R.M. Liskay, A. Letsou, J.L. Stachelek, Homology requirements for efficient gene conversion between duplicated chromosomal sequences in mammalian cells, Genetics 115 (1987) 161–167.

[44] A.S. Waldman, R.M. Liskay, Differential effect of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 5340–5344.

[45] A.S. Waldman, R.M. Liskay, Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology, Mol. Cell. Biol. 8 (1988) 5350–5357.

[46] D. Yang, A.S. Waldman, Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells, Mol. Cell. Biol. 17 (1997) 3614–3628.

[47] B.S. Lopez, P. Bertrand-Mercat, E. Corteggiani, J. Coppey, Directional recombination is initiated at a double strand break in human nuclear extracts, Nucleic Acids Res. 20 (1992) 501–506.

[48] J.A. Blaho, R.D. Wells, Left-handed Z-DNA and genetic recombination, Prog. Nucleics Acid. Res. Mol. Biol. 37 (1989) 107–126.

[49] W.P. Wahls, P.D. Moore, Homologous recombination enhancement conferred by the Z-DNA motif d(TG)30 is abrogated by Simian Virus 40 T antigen binding to adjacent DNA sequences, Mol. Cell. Biol. 10 (1990) 794–800.

[50] W.P. Wahls, L.J. Wallas, P.D. Moore, Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells, Cell 60 (1990) 95–103.

[51] R.G. Sargent, R.V. Merrihew, R. Nairn, G. Adair, M. Meuth, J.H. Wilson, The influence of a (GT)29 microsatellite se-

quence on homologous recombination in the hamster adenine phosphoribosyl transferase gene, Nucleic Acids Res. 24 (1996) 746–753.
[52] B.J. Thomas, R. Rothstein, Elevated recombination rates in transcriptionally active DNA, Cell 56 (1989) 619–630.
[53] K. Voelkel-Meiman, R.L. Keil, G.S. Roeder, Recombination-stimulating sequences in Yeast Ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I, Cell 48 (1987) 1071–1079.
[54] J.A. Nickoloff, R.J. Reynolds, Transcription stimulates homologous recombination in mammalian cells, Mol. Cell. Biol. 10 (1990) 4837–4845.
[55] J.A. Nickoloff, Transcription enhances intrachromosomal homologous recombination in mammalian cells, Mol. Cell. Biol. 12 (1992) 5311–5318.
[56] W.P. Deng, J.A. Nickoloff, Preferential repair of UV damage in highly transcribed DNA diminishes UV-induced intrachromosomal recombination in mammalian cells, Mol. Cell. Biol. 14 (1994) 391–399.
[57] D.G. Taghian, J.A. Nickoloff, Chromosomal double-strand breaks induce gene conversion at high frequency in mammalian cells, Mol. Cell. Biol. 17 (1997) 6386–6393.
[58] M.B. Benjamin, J.B. Little, X-rays induce interallelic homologous recombination at the human thymidine kinase gene, Mol. Cell. Biol. 12 (1992) 2730–2738.
[59] M.S. Park, Expression of human RAD52 confers resistance to ionizing radiation in mammalian cells, J. Biol. Chem. 270 (1995) 15467–15470.
[60] M. Brenneman, F.S. Gimble, J.H. Wilson, Stimulation of intrachromosomal homologous recombination in human cells by electroporation with site-specific endonuclease, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3608–3612.
[61] M. Jasin, Genetic manipulation of genome with rare-cutting endonucleases, TIG 12 (1996) 224–228.
[62] P. Rouet, M. Jasin, Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease, Mol. Cell. Biol. 14 (1994) 8096–8106.
[63] P. Rouet, F. Sih, M. Jasin, Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 6064–6068.
[64] G.R. Sargent, M.A. Brenneman, J.H. Wilson, Repair of site-specific double-strand breaks in a mammalian chromosome by homologous and illegitimate recombination, Mol. Cell. Biol. 17 (1997) 267–277.
[65] F. Liang, M. Han, P.J. Romanienko, M. Jasin, Homology-directed repair is a major double-strand break repair pathway in mammalian cells, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 5172–5177.
[66] C. Bonnerot, J.-F. Nicolas, Clonal analysis in the intact mouse embryo by intragenic homologous recombination, C. R. Acad. Sci. Paris 316 (1993) 1207–1217.
[67] J.R. Murti, M. Bumbulis, J.C. Schimenti, High-frequency germ line gene conversion in transgenic mice, Mol. Cell. Biol. 12 (1992) 2545–2552.
[68] F.-L. Lin, N. Sternberg, Homologous recombination between overlapping thymidine kinase gene fragments stably integrated into mouse cell genome, Mol. Cell. Biol. 4 (1984) 852–861.
[69] M.J. Shulman, C. Collins, A. Connor, L.R. Read, M.D. Baker, Interchromosomal recombination is suppressed in mammalian somatic cells, 14 (1995) 4102-4107.
[70] A.S. Waldman, B.C. Waldman, Stimulation of intrachrmosomal homologous recombination in mammalian cells by inhibitors of poly(ADP-ribosylation), Nucleic Acids Res. 19 (1991) 5943–5947.
[71] M.S. Park, D.L. Ludwig, E. Stigger, S.-H. Lee, Physical interaction between human RAD52 and RPA is required for homologous recombination in mammalian cells, J. Biol. Chem. 271 (1996) 18996–19000.
[72] S. Vispé, C. Cazaux, C. Lesca, M. Defais, Overexpression of Rad51 protein stimulates homologous recombination and increases resistance of mammalian cells to ionizing radiation, Nucleic Acids Res. 26 (1998) 2859–2864.